ABSTRACT

TANG, LIHUA. “Smooth” Inference for Clustered Survival Data. (Under the direction of Dr. Marie Davidian.)

Regression analysis of censored clustered-correlated time-to-event data is of interest in family studies, litter-matched tumorigenesis studies, and other settings where the survival times may be thought of as arising in groups or “clusters,” and the correlation among survival times in each cluster must be taken into account. A natural way to address such dependence is through incorporation of subject-specific random effects. In the first part of this dissertation, we propose an accelerated failure time (AFT) model for such data that involves normally-distributed, mean zero random effects and a within-cluster “error” term that is assumed to have distribution with a density satisfying mild “smoothness” conditions. We approximate the smooth density by the “seminonparametric” (SNP) representation of Gallant and Nychka (1987), which admits a “parametric” form for the density depending on a known “kernel” density and a tuning parameter that determines the degree of flexibility for capturing the true density. This representation facilitates likelihood-based inference on the regression parameter, random effects variance components, and the density, which we implement by a Monte Carlo expectation-maximization (MCEM) algorithm; and we choose the tuning parameter and “kernel” using standard information criteria. Moreover, arbitrary censoring patterns may be accommodated straightforwardly. We illustrate the approach via simulations and by applications to data from Diabetic Retinopathy Study (DRS, Diabetic Retinopathy Study Research Group, 1981), from a litter-matched tumorigenesis study (Mantel, Bohidar, and Ciminera, 1977), and from western Kenya parasitaemia study (McElroy et al., 1997).
The second part of this dissertation focuses on estimation of a bivariate survival function. In many situations, such as twin studies, matched pair studies, and studies of organ such as the eyes and kidneys, correlated, bivariate failure times are recorded. Based on a sample of possibly censored such failure times, an objective of analysis is to estimate the joint survival distribution. We extend the use of SNP in the first part of the dissertation to the two dimensional case and represent the joint density of the failure time using SNP. We illustrate the approach via simulations and by application to data from the DRS.
“Smooth” Inference for Clustered Survival Data

by

Lihua Tang

A dissertation submitted to the Graduate Faculty of North Carolina State University in partial fulfillment of the requirements for the Degree of Doctor of Philosophy

Statistics

Raleigh, North Carolina

2008

APPROVED BY:

______________________________  ______________________________
Marie Davidian                  Anastasios A. Tsiatis
Chair of Advisory Committee

______________________________  ______________________________
Daowen Zhang                    Wenbin Lu
Dedication

To my parents, my husband and my daughters
Biography

Lihua Tang was born on November 20, 1973 in Songyuan, Jilin province, P.R.China. She graduated from Nankai University with a bachelor’s degree in Mathematical Statistics in 1995. She entered the Department of Electrical Engineering at North Carolina State University to study Computer Engineering in 1999 and earned her master’s degree in Computer Engineering in 2000. After working at a software company for more than two years, she decided to join the Department of Statistics at North Carolina State University to study Statistics. She earned her master’s degree in Statistics in 2004 and continued her Ph.D studies in the Department.
Acknowledgements

I would like to express sincere gratitude to my advisor Dr. Marie Davidian for her constant guidance, encouragement, support and patience throughout the development of my doctoral research and the preparation of my thesis. I would also like to thank the other committee members, Dr. Anastasios A. Tsiatis, Dr. Daowen Zhang and Dr. Wenbin Lu for their insights and comments on my dissertation.

I gratefully acknowledge the faculties in Statistics at North Carolina State University for their high quality instruction and guidance. Special thanks go to Dr. John Monahan who provided a lot of Fortran help. I thank my friends, including my fellow graduate students and my supervisor at GlaxoSmithKline, Qiming Liao, for their assistance and friendship.

Finally, I would like to thank my parents and my in-laws for their support through this process. I especially want to thank my husband Junliang for his love, encouragement, support and patience, and my daughters Jessica and Rebecca who gave me so much pleasure.
Contents

List of Tables vii
List of Figures ix

1 Introduction 1
  1.1 Motivation ...................................................... 1
  1.2 Review of Clustered Correlated Time-to-Event Data Regression 3
    1.2.1 Proportional Hazards Model with Frailty .................. 3
    1.2.2 Accelerated Failure Time Model .......................... 5
  1.3 Review of Bivariate Survival Function Estimation ............... 9
    1.3.1 Bivariate Distribution Function .......................... 9
    1.3.2 Literature Review of Bivariate Survival Function Estimation 10
  1.4 Outline ......................................................... 15

2 Seminonparametric Density Representation 16
  2.1 Introduction ................................................... 16
  2.2 General SNP Density Representation .......................... 17
  2.3 Generating Samples from the SNP Density ...................... 23
    2.3.1 Acceptance-Rejection Method .............................. 23
    2.3.2 Sampling from a SNP Density .............................. 24

3 Accelerated Failure Time Regression Model for Clustered Correlated Data 29
  3.1 Accelerated Failure Time Regression Model ..................... 29
  3.2 SNP Representation of the Error Density ...................... 31
  3.3 Likelihood Function ............................................. 33
  3.4 Implementations .................................................. 35
    3.4.1 Monte Carlo EM Algorithm ................................ 35
    3.4.2 Rejection Sampling ........................................ 38
  3.5 Choosing $K$ and the “Kernel” Density ........................ 39
  3.6 Standard Errors and Confidence Intervals ........................ 40
List of Tables

4.1 Number of data sets choosing each $K$ and “kernel” combination. . . . 55
4.2 Right-censored data with normal error term and Uniform(0,1) covariate. 56
4.3 Right-censored data with normal error term and different Bernoulli(0.5) covariate in each cluster. Table entries are same as in Table 4.2. . . . 58
4.4 Right-censored data with normal error term and same Bernoulli(0.5) covariate in each cluster. Table entries are same as in Table 4.2. . . . 60
4.5 Right-censored data with mixture of normal error term and Uniform(0,1) covariate. Table entries are same as in Table 4.2. . . . . . . . . . . 62
4.6 Right-censored data with mixture of normal error term and different Bernoulli(0.5) covariate in each cluster. Table entries are same as in Table 4.2. . . . . . . . . . . . 64
4.7 Right-censored data with mixture of normal error term and same Bernoulli(0.5) covariate in each cluster. Table entries are same as in Table 4.2. . . . 66
4.8 Right-censored data with exponential error term and Uniform(0,1) covariate. Table entries are same as in Table 4.2. . . . . . . . . . . . 68
4.9 Right-censored data with exponential error term and different Bernoulli(0.5) covariate in each cluster. Table entries are same as in Table 4.2. . . . 70
4.10 Right-censored data with exponential error term and same Bernoulli(0.5) covariate in each cluster. Table entries are same as in Table 4.2. . . . 72
4.11 Right-censored data with Weibull error term and Uniform(0,1) covariate. Table entries are same as in Table 4.2. . . . . . . . . . . . 74
4.12 Right-censored data with Weibull error term and different Bernoulli(0.5) covariate in each cluster. Table entries are same as in Table 4.2. . . . 76
4.13 Right-censored data with Weibull error term and same Bernoulli(0.5) covariate in each cluster. Table entries are same as in Table 4.2. . . . 78
4.14 Right-censored data with Gamma error term and Uniform(0,1) covariate. Table entries are same as in Table 4.2 . . . . . . . . . . . . . 80
4.15 Interval censored data with mixture of normal error term and different Bernoulli(0.5) covariate in each cluster. Table entries are same in Table 4.2. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 82
4.16 Nominal level for variance components hypothesis testing

4.17 Right-censored data with mixture of normal error term, Uniform(0,1) covariate, and mixture of normal random effect. Table entries are same in Table 4.2.

4.18 Inference results for litter-matched tumorigenesis study data set

4.19 Inference results for DRS data set. Table entries are same as in Table 4.18.

4.20 Inference results for the Kenya parasitaemia data set. Table entries are the same as in Table 4.18.

6.1 Number of data sets choosing each $K$-base and “Kernel” combination for noncommon margins

6.2 NonCommon margins bivariate normal simulation results

6.3 Noncommon margins bivariate mixture normal simulation results. Table entries are same as in Table 6.2.

6.4 Noncommon margins bivariate mixture normal simulation results with bootstrap standard error.

6.5 Noncommon margins bivariate exponential simulation results. Table entries are same as in Table 6.2.

6.6 Noncommon margins bivariate Weibull simulation results. Table entries are same as in Table 6.2.

6.7 Number of data sets choosing each $K$-base and “Kernel” combination for common margins

6.8 Common margins bivariate normal simulation results. Table entries are same as in Table 6.2.

6.9 Common margins bivariate mixture of normal simulation results. Table entries are same as in Table 6.2.

6.10 Common margins bivariate exponential simulation results. Table entries are same as in Table 6.2.

6.11 Common margins bivariate Weibull simulation results. Table entries are same as in Table 6.2.

6.12 Hypothesis test for correlation

6.13 Inference for DRS data set
List of Figures

2.1 Histogram of 30000 random samples generated from a SNP density and true density ($K = 1, q = 1$) .................................................. 27

2.2 Histogram of 30000 random samples generated from a SNP density and true density ($K = 2, q = 1$) .................................................. 28

4.1 (a) Average of 100 estimated densities chosen by HQ (dash line) and true density of $e_{ij}$ (solid line). (b) 100 estimated densities chosen by HQ and true density (solid line): normal scenario with Uniform(0,1) covariate. ................................................................. 56

4.2 (a) Average of 100 estimated densities chosen by HQ (dash line) and true density of $\delta_{ij}$ (solid line). (b) 100 estimated densities chosen by HQ and true density (solid line): normal scenario with Uniform(0,1) covariate. ................................................................. 57

4.3 (a) Average of 100 estimated densities chosen by HQ (dash line) and true density of $e_{ij}$ (solid line). (b) 100 estimated densities chosen by HQ and true density (solid line): normal scenario with different Bernoulli(0.5) covariate in each cluster. .................................................. 58

4.4 (a) Average of 100 estimated densities chosen by HQ (dash line) and true density of $\delta_{ij}$ (solid line). (b) 100 estimated densities chosen by HQ and true density (solid line): normal scenario with different Bernoulli(0.5) covariate in each cluster. .................................................. 59

4.5 (a) Average of 100 estimated densities chosen by HQ (dash line) and true density of $e_{ij}$ (solid line). (b) 100 estimated densities chosen by HQ and true density (solid line): normal scenario with same Bernoulli(0.5) covariate in each cluster. .................................................. 60

4.6 (a) Average of 100 estimated densities chosen by HQ (dash line) and true density of $\delta_{ij}$ (solid line). (b) 100 estimated densities chosen by HQ and true density (solid line): normal scenario with same Bernoulli(0.5) covariate in each cluster. .................................................. 61
4.7 (a) Average of 100 estimated densities chosen by HQ (dash line) and true density of $e_{ij}$ (solid line). (b) 100 estimated densities chosen by HQ and true density (solid line): mixture of normal scenario with Uniform(0,1) covariate. ............... 62

4.8 (a) Average of 100 estimated densities chosen by HQ (dash line) and true density of $\delta_{ij}$ (solid line). (b) 100 estimated densities chosen by HQ and true density (solid line): mixture of normal scenario with Uniform(0,1) covariate. ......................... 63

4.9 (a) Average of 100 estimated densities chosen by HQ (dash line) and true density of $e_{ij}$ (solid line). (b) 100 estimated densities chosen by HQ and true density (solid line): mixture of normal scenario with different Bernoulli(0.5) covariate in each cluster. ............... 64

4.10 (a) Average of 100 estimated densities chosen by HQ (dash line) and true density of $\delta_{ij}$ (solid line). (b) 100 estimated densities chosen by HQ and true density (solid line): mixture of normal scenario with different Bernoulli(0.5) covariate in each cluster. ......................... 65

4.11 (a) Average of 100 estimated densities chosen by HQ (dash line) and true density of $e_{ij}$ (solid line). (b) 100 estimated densities chosen by HQ and true density (solid line): mixture of normal scenario with same Bernoulli(0.5) covariate in each cluster. ............... 66

4.12 (a) Average of 100 estimated densities chosen by HQ (dash line) and true density of $\delta_{ij}$ (solid line). (b) 100 estimated densities chosen by HQ and true density (solid line): mixture of normal scenario with same Bernoulli(0.5) covariate in each cluster. ......................... 67

4.13 (a) Average of 100 estimated densities chosen by HQ (dash line) and true density of $\delta_{ij}$ (solid line). (b) 100 estimated densities chosen by HQ and true density (solid line): exponential scenario with Uniform(0,1) covariate. ........................................ 68

4.14 (a) Average of 100 estimated densities chosen by HQ (dash line) and true density of $e_{ij}$ (solid line). (b) 100 estimated densities chosen by HQ and true density (solid line): exponential scenario with Uniform(0,1) covariate. ........................................ 69

4.15 (a) Average of 100 estimated densities chosen by HQ (dash line) and true density of $\delta_{ij}$ (solid line). (b) 100 estimated densities chosen by HQ and true density (solid line): exponential scenario with different Bernoulli(0.5) covariate in each cluster. ......................... 70

4.16 (a) Average of 100 estimated densities chosen by HQ (dash line) and true density of $e_{ij}$ (solid line). (b) 100 estimated densities chosen by HQ and true density (solid line): exponential scenario with different Bernoulli(0.5) covariate in each cluster. ......................... 71
4.17 (a) Average of 100 estimated densities chosen by HQ (dash line) and true density of $\delta_{ij}$ (solid line). (b) 100 estimated densities chosen by HQ and true density (solid line): Exponential scenario with same Bernoulli(0.5) covariate in each cluster. .............................. 72

4.18 (a) Average of 100 estimated densities chosen by HQ (dash line) and true density of $e_{ij}$ (solid line). (b) 100 estimated densities chosen by HQ and true density (solid line): Exponential scenario with same Bernoulli(0.5) covariate in each cluster. .............................. 73

4.19 (a) Average of 100 estimated densities chosen by HQ (dash line) and true density of $\delta_{ij}$ (solid line); (b) 100 estimated densities chosen by HQ and true density (solid line): Weibull scenario with Uniform(0,1) covariate. .......................................................... 74

4.20 (a) Average of 100 estimated densities chosen by HQ (dash line) and true density of $e_{ij}$ (solid line); (b) 100 estimated densities chosen by HQ and true density (solid line): Weibull scenario with Uniform(0,1) covariate. .......................................................... 75

4.21 (a) Average of 100 estimated densities chosen by HQ (dash line) and true density of $\delta_{ij}$ (solid line); (b) 100 estimated densities chosen by HQ and true density (solid line): Weibull scenario with different Bernoulli(0.5) covariate in each cluster. .............................................. 76

4.22 (a) Average of 100 estimated densities chosen by HQ (dash line) and true density of $e_{ij}$ (solid line); (b) 100 estimated densities chosen by HQ and true density (solid line): Weibull scenario with different Bernoulli(0.5) covariate in each cluster. .............................................. 77

4.23 (a) Average of 100 estimated densities chosen by HQ (dash line) and true density of $\delta_{ij}$ (solid line); (b) 100 estimated densities chosen by HQ and true density (solid line): Weibull scenario with same Bernoulli(0.5) covariate in each cluster. .............................................. 78

4.24 (a) Average of 100 estimated densities chosen by HQ (dash line) and true density of $e_{ij}$ (solid line); (b) 100 estimated densities chosen by HQ and true density (solid line): Weibull scenario with same Bernoulli(0.5) covariate in each cluster. .............................................. 79

4.25 (a) Average of 100 estimated densities chosen by HQ (dash line) and true density of $\delta_{ij}$ (solid line); (b) 100 estimated densities chosen by HQ and true density (solid line): Gamma scenario with Uniform(0,1) covariate. .......................................................... 80

4.26 (a) Average of 100 estimated densities chosen by HQ (dash line) and true density of $e_{ij}$ (solid line); (b) 100 estimated densities chosen by HQ and true density (solid line): Gamma scenario with Uniform(0,1) covariate. .......................................................... 81
4.27 (a) Average of 100 estimated densities chosen by HQ (dash line) and true density of $e_{ij}$ (solid line); (b) 100 estimated densities chosen by HQ and true density (solid line): Interval censored mixture of normal scenario with different Bernoulli(0.5) covariate. 82

4.28 (a) Average of 100 estimated densities chosen by HQ (dash line) and true density of $\delta_{ij}$ (solid line) and; (b) 100 estimated densities chosen by HQ and true density (solid line): Interval censored mixture of normal scenario with different Bernoulli(0.5) covariate. 83

4.29 (a) Average of 100 estimated densities chosen by HQ (dash line) and true density of $e_{ij}$ (solid line); (b) 100 estimated densities chosen by HQ and true density of $e_{ij}$ (solid line): Relaxing the distribution of random effect. 87

4.30 Dashed line is the density estimate based on normal assumption; solid line is the density estimate of our final chosen model. 89

4.31 (a) Histogram of failure time and estimate of density; (b) Survival function; (c) Hazard function. 90

5.1 SNP density representation with normal “kernel,” $K = 1$, and $q = 2$. 98

5.2 SNP density representation with normal “kernel,” $K = 2$, and $q = 2$. 99

5.3 SNP density representation with exponential “kernel” and $K = 1$. 101

6.1 (a) True bivariate normal density (b) Average of 100 density estimates 118

6.2 (a) True bivariate mixture normal density (b) Average of 100 density estimates 119

6.3 (a) True bivariate exponential density (b) Average of 100 density estimates 121

6.4 (a) True bivariate Weibull density (b) Average of 100 density estimates 122

6.5 (a) True bivariate normal density (b) Average of 100 density estimates 126

6.6 (a) True bivariate mixture normal density (b) Average of 100 density estimates 128

6.7 (a) True bivariate exponential density (b) Average of 100 density estimates 130

6.8 (a) True bivariate weibull density (b) Average of 100 density estimates 132

6.9 Estimated bivariate density for DRS data 137

6.10 (a) Estimated marginal density of $T_{1l}$; (b) Estimated marginal density of $T_{2l}$. 138

6.11 Estimated bivariate survival function $P_r(T_1 \geq t_1, T_2 \geq t_2)$ for DRS data 139

6.12 Estimated bivariate survival function $P_r(T_1 \geq t, T_2 \geq t)$ for DRS data 140
Chapter 1

Introduction

1.1 Motivation

In many family studies, multi-center clinical trials, litter-matched tumorigenesis studies, and other settings, the outcome of interest may be a possibly censored time to an event (survival time or failure time), and the survival times may be thought of as arising in groups or “clusters.” Also, proper assessment of prognostic factors requires acknowledging heterogeneity across families, centers, or litters, which may be accommodated by viewing the data as “clustered.” For example, in a litter-matched tumorigenesis study (Mantel et al., 1977), there were three rats in each of 50 female litters. In each litter, one rat was drug-treated and another two were controls. The failure time of interest is time to tumor appearance, which may be right-censored, where the censoring time is time to death. Here, litter effects may be thought of “clusters,” and there might be correlation among the survival times in each litter.
We refer to such data as correlated clustered time-to-event data. The investigators are interested in assessing whether the time to tumor appearance for the control group tends to be longer than in the treatment group. Another interest is in testing whether or not the times to tumor appearance are indeed correlated in each litter.

In general, a characteristic of correlated clustered time-to-event data is that the individuals (subjects, rats, etc.) within each cluster share common genetic or environmental factors such that the failure times within each cluster might be correlated. Dependence induced by clustering must be taken into appropriate account in order to obtain valid inferences on questions of interest. If the failure times represent the same type of time-to-event, we refer to them as “parallel” event data. Recurrent event times, such as depressive episodes from the same subject, present another case of clustered correlated survival times. Many multivariate survival analysis models (Lin et al., 2000; Pena, Strawderman, and Hollander, 2001; Huang and Wang, 2005) have been proposed to deal with recurrent failure data. Here, the “cluster” is the subject on whom recurrent events are observed. The first part of this dissertation focuses on new methods for regression analysis of clustered correlated time-to-event data.

In many situations, such as twin studies, matched pair studies, and studies of organ such as the eyes and kidneys, correlated, bivariate failure times are recorded. Based on a sample of possibly censored such clustered failure times, an objective of analysis is to make inference on the joint survival distribution. Correlation between the two failure times must be taken into account. In the second part of this dissertation, we consider new methods for this type of analysis.
In the methods we develop for regression analysis of clustered correlated survival data, data may be censored according to some arbitrary censoring pattern. For the methods we propose for estimation of a bivariate survival function, data are only considered in the presence of right censoring, but the methods may be extended to other censoring patterns. Our approach is to assume that the distribution of survival times has a “smooth” density. In the remainder of this section, we review previous literature on these problems as background, and we introduce what we mean by a “smooth” density in Chapter 2.

1.2 Review of Clustered Correlated Time-to-Event Data Regression

There is an extensive literature on regression modeling of clustered correlated time-to-event data. Two popular models are reviewed in this section. Section 1.2.1 reviews the proportional hazards model with frailty; and Section 1.2.2 reviews methods based on the accelerated failure time model.

1.2.1 Proportional Hazards Model with Frailty

One approach to modeling clustered correlated time-to-event data is to extend the Cox proportional hazards (PH) model (Cox, 1972) by introducing a cluster-specific random effect, a so-called “frailty” (Hougaard, 2000; Therneau and Grambsch, 2000). The frailty might be viewed as the random effect shared by individuals in each cluster,
and this frailty induces the dependence among the failure times.

Let \( n \) denote the number of clusters; \( X_{ij} \) be covariates; and \( T_{ij} \) be the \( j \)th failure time in cluster \( i, i = 1, \ldots, n, j = 1, \ldots, m_i \). Let \( \omega_i, i = 1, \ldots, n, \) be a cluster-specific frailty. Defining the hazard function of the failure time given the covariate and the frailty as usual by

\[
\lambda(t|X_{ij},\omega_i;\beta) = \lim_{\delta \to 0^+} \frac{\delta^{-1}P_r(t \leq T_{ij} < t + \delta|T_{ij} \geq t, X_{ij}, \omega_i; \beta)}{\delta},
\]

then the proportional hazard model with frailty may be represented as

\[
\lambda(t|X_{ij},\omega_i;\beta) = \omega_i \lambda_0(t) \exp(\beta^T X_{ij}),
\]

where \( \lambda_0(x) \) is an arbitrary baseline hazard function, and \( \beta \) is an unknown \((p \times 1)\) vector of regression parameters. The frailty is often assumed to have some parametric distribution, such as the log-normal or gamma distribution. Clayton and Cuzick (1985) assumed that the frailty comes from a gamma distribution with mean one and unknown variance. They estimated the frailty parameters and covariate effects by using a modified EM algorithm. McGilchrist and Aisbett (1991) considered a lognormal distribution for frailty. Klein (1992) proposed an EM algorithm to estimate the frailty parameter and covariate effects and also assumed that the frailty arises from a gamma distribution. Murphy (1995) proved the asymptotic properties for the nonparametric likelihood estimator in the gamma frailty proportional hazard model without covariates. Hougaard (1986) and Fine, Glidden, and Lee (2003) considered the positive stable distribution for the frailty distribution. Therneau, Grambsch and Pankratz (2003) proposed a penalized likelihood method to estimate the frailty parameters and covariate effects in the lognormal and \( t \) frailty models. Much effort has been devoted
to relaxing the assumptions on the distribution of the frailty (Hsu, Gorfine, and Malone, 2007). However, the choice of frailty distribution may have an important impact on estimating the regression parameters (Hougaard, 2000, Chap. 7). Also, the PH model is generally not robust to failure to include important covariates (Hougaard, 1999).

1.2.2 Accelerated Failure Time Model

Another alternative to the PH model is the accelerated failure time (AFT) model (Kalbfleisch and Prentice, 2002, Section 2.2.3), which assumes that the covariates can either speed up or slow down the expected failure time. The advantage of the AFT model is that it models the failure time directly as a function of covariates, while the PH model accomplishes this indirectly through the hazard function.

Again, let $T_{ij}$ be the $j$th failure time in cluster $i$, and let $X_{ij}$ be a corresponding vector of covariates. For correlated clustered survival data, one approach to address the within-cluster dependence is through the marginal AFT model, given by

$$\log(T_{ij}) = X_{ij}^T \beta + \epsilon_{ij},$$

where $\beta = (\beta_1, \ldots, \beta_p)^T$ is an unknown vector of regression parameter, and $\epsilon_{ij}$ is an “error” term. Let $\epsilon_i = (\epsilon_{i1}, \ldots, \epsilon_{im_i})^T$, $i = 1, \ldots, n$, for clustered data, it is unlikely that the $\epsilon_{ij}$ will be uncorrelated for given $i$.

Lin and Wei (1992) proposed a semiparametric method for fitting (1.1) in which the distribution form of error term $\epsilon_{ij}$ is not specified in the analysis. The method is based on rank statistics and Buckley-James (Buckley and James, 1979) least squares
estimation procedure, and can be applied to the case of two or more distinct failure
times in each cluster. Lin, Wei, and Ying (1998) proposed estimation based on ranks.
However, it is very difficult to calculate the rank estimators because the estimating
functions are step functions that may have multiple roots, and it is difficult to identify
a consistent root. Estimation of the variance-covariance matrix is also a further
difficulty. Jin et al. (2003) proposed a class of monotone estimating functions that
approximates the weighted log-rank estimating functions around the true regression
parameters for univariate failure time data; Jin, Lin, and Ying (2006) extended this
approach to multivariate failure time data.

Lee, Wei, and Ying (1993) proposed a popular semi-parametric regression method
called the marginal independence approach for fitting (1.1). The marginal indepen-
dence approach estimates the regression coefficient by ignoring the correlation among
the failure times. The drawback of this method is the obvious efficiency loss. Pan
(2001) extended Lee et al. (1993), incorporating a frailty in the accelerated failure
time model to explicitly account for possible correlations among the failure times in
each cluster. They assumed that the hazard function of the within-cluster error term
$\epsilon_{ij}$, given a random frailty $\omega_i$, $i = 1, \ldots, n$, can be represented by $\omega_i$ times some
baseline hazard function, which is independent of the covariate, i.e.,

$$h(\epsilon_{ij}|\omega_i) = \omega_i h_0(\epsilon_{ij}),$$

where $h_0(\epsilon_{ij})$ is an arbitrary baseline hazard function; $\omega_1, \ldots, \omega_n$ are independent
identical distributed (i.i.d.) and are not dependent on $X_{ij}$. The dependence among
the failure times in the cluster is induced by the random frailty $\omega_i$, such that the
error terms $\epsilon_{i1}, \ldots, \epsilon_{im_i}$ are correlated within a cluster. They assumed that $\omega_i$'s are
an i.i.d sample from a gamma distribution $G(\theta, 1/\theta)$ with shape and inverse scale parameter both $\theta$. An EM-like algorithm was adopted to obtain the likelihood estimates. However, the algorithm can suffer numerical unstablity with respect to $\theta$. They followed the profile likelihood (Nielsen et al., 1992). For a given $\theta \in \Theta$, an EM-like algorithm was used to estimate other parameters, and the final estimate for each parameter was obtained by maximizing observed loglikelihood. In Pan’s (2001) paper, if all the estimates of $\omega_i$ are equal to 1, then, there is no frailty. The score function is reduced to the estimating equation of the marginal independence model in Lee et al. (1993), which is same as in the univariate case (Tsiatis, 1990). However, his estimation method is unstable because of the discontinuity of its estimating equation like other nonparametric estimation methods. Therefore, Zhang and Peng (2007) proposed another nonparametric estimation method for the marginal AFT model assuming (1.1) and used the EM algorithm to solve the estimating equations. They applied an M-estimator in the M-step to simplify the estimation equations. However, one issue for this proposed method is the discontinuity of the estimating equations and another issue is that the estimator of the variance of the estimates is not efficient because of the complexity of the estimating equations.

Another approach to address within-cluster dependence is through incorporation of subject-specific random effects directly in the AFT model. The general subject-specific AFT model is given by

$$\log(T_{ij}) = X_{ij}^T \beta + S_{ij}^T b_i + e_{ij},$$

(1.2)

where $X_{ij}$ is a vector of covariates; $\beta = (\beta_1, \cdots, \beta_p)^T$ is an unknown regression coefficient vector; $b_i = (b_{i1}, \cdots, b_{il})^T$ is a random effect vector; $S_{ij}$ is the covariate
vector for random effects; and \( e_{ij} \) is a within-cluster error term. The \( b_i \) and \( e_{ij} \) are each assumed i.i.d. and independent of each other. Let \( N = \sum_{i=1}^{n} m_i \). Model (1.2), restricted to \( S_{ij} \equiv 1 \), which only has a random intercept, was studied by Pan and Louis (2000) and Pan and Connett (2001) for right-censored data. They proposed an EM-like algorithm, where censored survival time are imputed through a Buckley-James type approach. But they required that the random effect and error term have common marginal distribution, approximated using the Kaplan-Meier estimator. This difficulty stems in part from their focus on being “non-parametric” with respect to the marginal survival distribution. Therneau et al. (2003) proposed a penalized method to estimate the unknown parameters in the subject-specific AFT model (1.2) by treating the frailty as additional regression coefficients. Lambert, Kimber, and Johnson (2004) used a parametric method to estimate all the unknown parameters. Komárek and Lesaffre (2007) proposed a Bayesian approach for the subject-specific AFT model (1.2) with interval censored data. A Markov Chain Monte Carlo (MCMC) algorithm was developed to fit the model, in which the random effects are assumed normally distributed and i.i.d. and the \( e_{ij} \) are assumed to arise from a distribution with a density that can be represented by a finite mixture of normal densities, where the number of components in the mixture is chosen by the MCMC algorithm. This formulation allows flexibility in the assumptions on the error term.

In Chapter 3 and 4 of this dissertation, we consider model (1.2), restricting to \( S_{ij} \equiv 1 \) and make assumptions in a similar spirit to those of Komárek and Lesaffre (2007). We assume as they do normal random effects; however, we instead follow Zhang and Davidian (2007) and use a different representation of the density of the
error term, approximating it by the so-called “seminonparametric” (SNP) density representation of Gallant and Nychka (1987). The SNP density representation is reviewed in Chapter 2. The SNP representation facilitates the likelihood-based inference, which we implement by a Monte Carlo expectation-maximization (MCEM) algorithm; details are in Chapter 3.

1.3 Review of Bivariate Survival Function Estimation

Before we discuss existing approaches to estimating a bivariate survival function, we review the definition and properties of a bivariate distribution function.

1.3.1 Bivariate Distribution Function

Let $F_1(t_1)$ and $F_2(t_2)$, $f_1(t_1)$ and $f_2(t_2)$, $S_1(t_1)$ and $S_2(t_2)$ be the cumulative distribution functions (cdfs), densities, and survival functions of continuous random variables $T_1$ and $T_2$, where $S_k(t) = 1 - F_k(t) = P_r(T_k > t), k = 1, 2$. Then a bivariate probability function $F(t_1, t_2)$ with these marginal distributions is monotonically increasing from zero to unity and is subject to the following conditions:

a) $F(-\infty, t_2) = F(t_1, -\infty) = 0$,

\[ F(t_1, \infty) = F_1(t_1); \quad F(\infty, t_2) = F_2(t_2); \quad F(\infty, \infty) = 1. \]

b) For every $t_{11} < t_{12}, t_{21} < t_{22}$,
\begin{align*}
P_r(t_{11} < T_1 \leq t_{12}, t_{21} < T_2 \leq t_{22})
&= F(t_{12}, t_{22}) - F(t_{12}, t_{21}) - F(t_{11}, t_{22}) + F(t_{11}, t_{21}) \geq 0.
\end{align*}

If the second cross partial derivative \( \frac{\partial^2 F(t_1, t_2)}{\partial t_1 \partial t_2} \) exists everywhere, then the bivariate distribution has a density \( f(t_1, t_2) \) given by

\[
\frac{\partial^2 F(t_1, t_2)}{\partial t_1 \partial t_2} = f(t_1, t_2) \geq 0,
\]

and the joint survival function \( S(t_1, t_2) = P_r(T_1 \geq t_1, T_2 \geq t_2) \), which can be further written as

\[
S(t_1, t_2) = \int_{t_1}^{\infty} \int_{t_2}^{\infty} f(x, y)dxdy.
\]

### 1.3.2 Literature Review of Bivariate Survival Function Estimation

The Diabetic Retinopathy Study (DRS, Diabetic Retinopathy Study Research Group, 1981) was a randomized trial conducted by the National Eye Institute to evaluate the photocoagulation treatment for proliferative diabetic retinopathy. Patients with diabetic retinopathy and visual acuity of 20/100 or better in both eyes were eligible for the study. One eye of each patient was randomly chosen for photocoagulation treatment and the other eye was untreated. The failure time is the occurrence of visual acuity less than 5/200 at two consecutively completed 4-month follow-ups. The study involved 1,742 patients. A subset of the study with a sample of the high-risk patients as defined by DRS criteria is analyzed (n = 197). In such a study, interest may focus on estimation of joint bivariate survival function \( S(t_1, t_2) \),
where $T_1$ corresponds to failure time for an treated eye and $T_2$ to failure time for a untreated eye. The censoring variable was not described in the original study report. Each eye was censored separately, so that there are two distinct censoring variables, $C_1$ and $C_2$, corresponding to potential censoring times for $T_1$ and $T_2$, respectively.

Various approaches have been proposed to estimate the bivariate survival function. Hougaard (2000, Chap. 1) distinguishes among forms of multivariate survival data; here, we consider “parallel” data, where two study units (e.g., eyes, lungs, or kidneys) are followed simultaneously or where two different endpoints are monitored on the same subject. Censoring may be “univariate” (Lin and Ying, 1993), where both event times are censored simultaneously, or involve more general mechanisms and hence greater methodological challenges.

There is an extensive literature on non-parametric estimation of the bivariate survival function (e.g., Tsai, Leurgans, and Crowley, 1986; Dabrowska, 1988; Prentice and Cai, 1992; Pruitt, 1991; van der Laan, 1996; Quale, van der Laan, and Robins, 2006). It is well-known that the bivariate nonparametric maximum likelihood estimator (NPMLE) for continuous data is not consistent (Tsai et al., 1986). Pruitt (1991) proposed an estimator that can solve the non-uniqueness of the NPMLE by estimating conditional densities over the half lines implied by the singly censored observations (Quale, van der Laan, and Robins, 2006). However, this estimator is not asymptotically efficient. Dabrowska (1988) and Prentice and Cai (1992) proposed a “plug-in” nonparametric estimator. The estimator expresses the bivariate survival function as a product of the marginal survival function and a dependence function. Both use Kaplan-Meier estimates for the marginal survival functions but they es-
timate the dependence function differently. Both proposed estimators are strongly consistent, asymptotically normal, and nonparametric efficient under complete independence of failure time and right-censoring time. Both methods perform well in relatively small sample sizes and in the presence of heavy censoring and are locally efficient in the sense that both are efficient when the failure time and censoring are jointly independent (Gill, van der Laan, and Wellner, 1995). Dabrowska’s estimator, Prentice and Cai’s estimator, and van der Laan’s (1996) estimator have better performance in finite samples (van der Laan, 1997). van der Laan (1992, 1996) developed a sequence of reductions of NPMLE (SOR-NPMLE) estimator, which is based on reduced data, where he reduced the uncensored component of the singly censored observations into interval censored data by using a lattice partition. van der Laan proved that the SOR-NPMLE is uniformly consistent and asymptotically normal. Quale, van der Laan, and Robins (2006) proposed a local efficient estimator which has a double-robust property.

Another alternative is to model the bivariate distributions using copula models; see, for example, Plackett (1965), Clayton (1978), Cook and Johnson (1981), Genest and Mackay (1986), Hougaard (1986), Marshall and Olkin (1988), Nelsen (1986), and Shih and Louis (1995). Suppose that $C_\alpha$ is a distribution function with density $c_\alpha$ on $[0, 1]^2$ for $\alpha \in R^1$. Then the joint survival function and density function of $(T_1, T_2)$ is given by

$$S(t_1, t_2) = C_\alpha\{S_1(t_1), S_2(t_2)\} \quad t_1, t_2 \geq 0$$

$$f(t_1, t_2) = c_\alpha\{S_1(t_1), S_2(t_2)\} f_1(t_1)f_2(t_2) \quad t_1, t_2 \geq 0$$

One attractive feature of a copula model is that the margins do not depend on the
choice of the dependence structure and the dependency and the margins can be estimated separately. Therefore, two-stage estimation is often used. Stage I estimates the two margins assuming independence and Stage II estimates the association parameter by fixing the margins at the estimate from Stage I. The dependence was induced by the association parameter $\alpha$. In order to make inferences about $\alpha$, we may parameterize the marginal survival functions of $T_1$, $G(t_1) = G(t_1, \tau)$ and of $T_2$, $H(t_2) = H(t_2, \eta)$ and regard $\tau$ and $\eta$ as nuisance parameters. Oakes (1982) discussed the parametric inference for the case of $G$ and $H$ both exponential (Oakes, 1986). Considerate literature suggested parametric methods (Clayton, 1978; Oakes, 1982, 1986; Shih and Louis, 1995). Clayton (1978) suggested a pseudolikelihood estimation, where the pseudolikelihood is a product of the conditional probabilities. Oakes (1982) proposed a simpler estimator based on Kendall’s coefficient of concordance and derived the asymptotic variance. In Clayton-Oakes models, any form of the marginal distribution may be used, $\alpha$ can be interpreted as a relative risk, the risk of failing for member one if the second member fails relative to the risk of failing for member one if the second member does not fail. The disadvantage is that the association modeled must be positive. Genest and Mackay (1986) extended this approach to nonpositive association. Clayton and Cuzick (1985) also represented Clayton’s pseudolikelihood estimator as a weighted form of Oakes’s concordance estimator. Shih and Louis (1995) proposed a two-stage parametric estimation method, assuming the functional form of the margins is known.

A semiparametric approach is another popular method to estimate the bivariate survival function. Genest and Mackay (1986) introduced Archimedean copulas, which
is subclass of copulas characterized by some generator. Inference does not depend on
the form of the marginal distributions, which are treated as nuisance functions. Shih
and Louis (1995) also proposed a two-stage semiparametric method. At stage I, they
relaxed the parametric assumptions on the two margins and estimated the survival
functions by Kaplan-Meier estimates. They derived the asymptotic properties of
the estimators under some regularity conditions. Ghosh (2006) extended the second
model in Clayton (1978) and estimated the association parameter in the Plackett
distribution by the estimation method described in Shih and Louis (1995). Neato-
Barajas et al. (2007) proposed a Bayesian semiparametric method. The marginal
densities are well known nonparametric survival models, and the joint density is
constructed via a mixture. They defined the margins by hazard function and also
defined a new copula. Another alternative to estimate the bivariate survival function
is based on kernel estimation. Wells and Yeo (1996) proposed kernel estimation,
including bandwidth selection and boundary effects. Kooperberg (1998) used linear
splines and their tensor products. Lambert (2007) proposed Archimedean copula
estimation using Bayesian splines smoothing techniques.

In Chapters 5 and 6 of this dissertation, we propose a new approach to representing
and estimating a bivariate survival function in the presence of right censoring. As
for the error term in the AFT model (1.2) of Section 1.2, we use the SNP density
representation in two dimensions to approximate the form of the survival function
and develop likelihood-based inference; details are in Chapter 5.
1.4 Outline

The theme of this dissertation is that, by making mild “smoothness” assumptions on components of the two types of models considered in Section 1.2 and 1.3, one can develop practically feasible methods that circumvent some of the problems associated with completely nonparametric methods and gain efficiency.

In Chapter 2, we introduce the SNP density representation that forms the basis for our proposed “smooth” approaches to both problems considered in Section 1.2 and 1.3. In Chapter 3, we describe the cluster-specific accelerated failure time model in which the error term is assumed to have a “smooth” density that can be represented by SNP, and give the proposed Monte Carlo EM algorithm to implement likelihood-based inference including the strategy to choose the degree of flexibility required for the SNP density. We also present the likelihood ratio tests for the variance components. We illustrate the methods by applications to three data sets and present simulations demonstrating performance of the approach in Chapter 4. In Chapter 5, we describe the method to estimate bivariate survival function under the “smoothness” condition. We illustrate the methods by application to data from the DRS study and present simulations demonstrating performance of the approach in Chapter 6.
Chapter 2

Seminonparametric Density Representation

2.1 Introduction

The approaches to regression analysis of clustered correlated survival data and bivariate survival function estimation proposed in subsequent chapters rely on approximating the assumed density of the within-cluster “error” term in the first model and the bivariate survival function in the second model by the so-called “seminonparametric” (SNP) density representation of Gallant and Nychka (1987). In this Chapter, we review this representation and related topics.

The term “seminonparametric” was coined by Gallant and Nychka (1987) to describe an approach that lies halfway between fully nonparametric and completely
parametric forms. Gallant and Nychka (1987) demonstrated that “smooth” densities may be represented by an infinite Hermite series expansion plus a lower bound on tail behavior, where “smooth” means that the density satisfies certain smoothness restrictions, including differentiability conditions. A full description of “smoothness” was given by Gallant and Nychka (1987); the class of such densities includes densities that are multi-modal, fat- or thin-tailed relative to the normal or exponential densities or skewed. They suggested that a truncated version of an infinite Hermite series expansion could be used as an approximation to a “smooth” density in practice.

We give a detailed description of the general SNP representation and its implementation in Section 2.2. An acceptance-rejection algorithm to generate a random sample from a SNP density with an exponential “kernel” is addressed in Section 2.3.

2.2 General SNP Density Representation

Suppose $Z$ is a $q$–variate random vector with a density $h(z) \in \mathcal{H}$, where $\mathcal{H}$ is a certain class of “smooth” densities. Gallant and Nychka (1987) showed that $h(z)$ may be approximated by a truncated Hermite expansion $h_K(z)$,

$$h_K(z) = P_K^2(z)\xi_q(z),$$

(2.1)

for some fixed value of $K$, where $P_K(z)$ is a truncated multivariate polynomial of degree $K$, and $\xi_q(\cdot)$ is a known “kernel” density with a moment generating function. In subsequent chapters, $\xi_q(\cdot)$ is taken to be either the density function of the standard $q$–variate normal distribution $N(0, I_q)$, which has mean 0 and variance–covariance matrix $I_q$ (referred to as the normal “kernel”) or the density function of the standard
q-variate exponential distribution with the standard exponential distribution margins and variance–covariance matrix \( I_q \) (the exponential “kernel”).

The truncated polynomial of degree \( K \) is given by \( P_K(z) = \sum_{|\lambda| \leq K} a_\lambda z^\lambda \), where \( \lambda = (\lambda_1, \ldots, \lambda_q)^T \) is a \( q \)-dimension vector of nonnegative integers, \( z^\lambda \) is the monomial \( z_1^{\lambda_1} \cdots z_q^{\lambda_q} \) of order \( |\lambda| = \sum_{k=1}^q \lambda_k \). For a particular choice of \( q \), there will be \( d \) distinct terms in the truncated polynomial; e.g., for \( q = 1 \), \( d = K + 1 \); for \( q = 2 \), \( d = (K + 1)(K + 2)/2 \). The coefficients \( a = (a_1, \ldots, a_d)^T \) are chosen so that

\[
\int h_K(z) dz = 1. \tag{2.2}
\]

As shown by Zhang and Davidian (2001), (2.2) is equivalent to requiring that

\[
\mathbb{E}\{P_K^2(U)\} = 1, \tag{2.3}
\]

where \( U \) is a \( q \)-variate random vector with density \( \xi_q(u) \). Zhang and Davidian (2001) found that numerical instability in estimating the polynomial coefficients may result if the SNP representation (2.1) given above is used directly. They suggested using a polar coordinate transformation, which can provide a unique SNP representation.

To have a convenient way to traverse the set \( \{ \lambda : 0 \leq |\lambda| \leq K \} \), let the \( d \) elements of \( \{ \lambda : 0 \leq |\lambda| \leq K \} \) be ordered in some arbitrary way and denote \( a_\lambda = a_i, U^\lambda = V_i, \) \( i = 1, \ldots, d \). Then

\[
P_K(U) = \sum_{i=1}^d a_i V_i = a^T V,
\]

where \( V = (V_1, \ldots, V_d)^T \).

For example, with \( q = 1, K = 2 \), \( a = (a_1, a_2, a_3)^T \), and \( V = (1, U, U^2)^T \). Then

\[
P_2(U) = a_1 + a_2 U + a_3 U^2.
\]
With \( q = 2 \) and \( K = 2 \), so that \( U = (U_1, U_2)^T \), let \( a = (a_1, a_2, a_3, a_4, a_5, a_6)^T \), and \( V = (1, U_2, U_2^2, U_1, U_1^2, U_1 U_2)^T \). Then
\[
P_2(U) = a_{00} + a_{10} U_1 + a_{01} U_2 + a_{20} U_1^2 + a_{02} U_2^2 + a_{11} U_1 U_2.
\]
Therefore,
\[
E\{P^2_K(U)\} = E(a^T V V^T a) = a^T E(V V^T) a = a^T A a = 1,
\]
where \( A = E(V V^T) \). It is well known that \( A \) is a positive definite matrix. Thus it has a decomposition
\[
A = B^T B,
\]
where \( B \) is an upper triangular matrix. Substituting (2.5) and (2.4) into (2.3), we obtain \( a^T B^T B a = 1 \). Thus letting \( c = B a \), we have \( c^T c = 1 \). This suggests a polar coordinate transformation for the reparameterization, given by
\[
c_1 = \sin(\psi_1)
\]
\[
c_2 = \cos(\psi_1) \sin(\psi_2)
\]
\[
\vdots
\]
\[
c_j = \cos(\psi_1) \cos(\psi_2) \cdots \sin(\psi_j)
\]
\[
\vdots
\]
\[
c_{d-1} = \cos(\psi_1) \cos(\psi_2) \cdots \sin(\psi_{d-1})
\]
\[
c_d = \cos(\psi_1) \cos(\psi_2) \cdots \cos(\psi_{d-1}),
\]
where \( c = (c_1, \ldots, c_d)^T \) and \( \psi = (\psi_1, \ldots, \psi_{d-1})^T \). To achieve a unique representation, we require \( \psi_j \in (-\pi/2, \pi/2] \), \( j = 1, \ldots, d - 1 \). Thus, with \( c = c(\psi) \),
\[
a = B^{-1} c(\psi),
\]
19
where $c(\psi)$ is a function of $\psi$ satisfying (2.6). Therefore, the SNP density can be represented in terms of $\psi$ satisfying the condition $E\{P_K^2(U)\} = 1$. The advantage of this reparameterization is that it is more stable to estimate the parameter $\psi$ instead of $a$. From now on, we often write $P_K(z)$ as $P_K(z, \psi)$ or $P_K(z, a)$ to emphasize dependence on $\psi$ or $a$.

For example, with $K = 2, q = 1, a = (a_1, a_2, a_3)^T, V = (1, U, U^2)^T$ and $U \sim N(0, 1)$. Then

$$P_2(U; a) = a_1 + a_2 U + a_3 U^2.$$ 

Thus

$$A = E(V V^T) = E \begin{bmatrix} 1 & U & U^2 \\ U & U^2 & U^3 \\ U^2 & U^3 & U^4 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 3 \end{bmatrix},$$

$$B = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & \sqrt{2} \end{bmatrix},$$

$$B^{-1} = \begin{bmatrix} 1 & 0 & -1/\sqrt{2} \\ 0 & 1 & 0 \\ 0 & 0 & 1/\sqrt{2} \end{bmatrix},$$

and

$$c(\psi) = \begin{bmatrix} \sin \psi_1 \\ \cos \psi_1 \sin \psi_2 \\ \cos \psi_1 \cos \psi_2 \end{bmatrix}.$$
Therefore, the coefficients $a$ in $P_2(U; a)$ are given by

$$a = B^{-1}c(\psi) = \begin{bmatrix}
\sin \psi_1 - \cos \psi_1 \cos \psi_2 / \sqrt{2} \\
\cos \psi_1 \sin \psi_2 \\
\cos \psi_1 \cos \psi_2 / \sqrt{2}
\end{bmatrix}.$$ 

With $K = 1, q = 1, a = (a_1, a_2)^T, V = (1, U)^T$ and $U \sim \text{Exp}(1)$, where $\text{Exp}(1)$ is the exponential density with mean 1. Then

$$P_1(U; a) = a_1 + a_2 U.$$ 

Thus

$$A = E(VV^T) = E\begin{bmatrix} 1 & U \\ U & U^2 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix},$$

$$B = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix},$$

$$B^{-1} = \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix},$$

and

$$c(\psi) = \begin{bmatrix} \sin \psi_1 \\ \cos \psi_1 \end{bmatrix}.$$ 

Therefore, the coefficients $a$ in $P_1(U; a)$ are given by

$$a = B^{-1}c(\psi) = \begin{bmatrix} \sin \psi_1 - \cos \psi_1 \\ \cos \psi_1 \end{bmatrix}.$$ 

Here, we give one more example for two dimensional case, which is needed to represent bivariate density in Chapter 5. With $K = 1, q = 2, a = (a_1, a_2, a_3)^T, V =$
$(1, U_2, U_1)^T$ and $U_1, U_2 \sim \text{Exp}(1)$ independently. Then

$$P_1(U; a) = a_1 + a_3 U_1 + a_2 U_2$$

Thus

$$A = E(VV^T) = E\begin{bmatrix} 1 & U_2 & U_1 \\ U_2 & U_2^2 & U_1 U_2 \\ U_1 & U_1 U_2 & U_1^2 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{bmatrix},$$

$$B = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix},$$

$$B^{-1} = \begin{bmatrix} 1 & -1 & -1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix},$$

and

$$c(\psi) = \begin{bmatrix} \sin \psi_1 \\ \cos \psi_1 \sin \psi_2 \\ \cos \psi_1 \cos \psi_2 \end{bmatrix}.$$ 

Therefore, the coefficients $a$ in $P_1(U; a)$ are given by

$$a = B^{-1} c(\psi) = \begin{bmatrix} \sin \psi_1 - \cos \psi_1 \sin \psi_2 - \cos \psi_1 \cos \psi_2 \\ \cos \psi_1 \sin \psi_2 \\ \cos \psi_1 \cos \psi_2 \end{bmatrix}.$$ 

To use this representation in practice, one must choose the degree of truncation $K$ and the “kernel” density $\xi_q(\cdot)$. In most use of SNP in the literature, the “kernel” is taken to be the $N(0, I_q)$ density (Zhang and Davidian, 2001; Chen et al., 2002). More
recently, Zhang and Davidian (2007) and Doehler and Davidian (2008) considered choosing between a $N(0, I_q)$ and a standard exponential “kernel” in order to achieve greater flexibility for representing densities that would be likely densities for failure time random variables. In all of this work, the choice of $K$ (and “kernel”) is usually made by inspection of standard information criteria. In subsequent chapters, we describe in detail the use of this approach in the proposed methods.

2.3 Generating Samples from the SNP Density

An advantage of the SNP density representation is that, once the representation is chosen, it is straightforward to simulate samples from it. This allows the user to approximate any functional of the density by Monte Carlo simulation. In later chapters, we demonstrate the utility of this capability with our methods. In this section, we discuss how such samples are generated.

2.3.1 Acceptance-Rejection Method

Kennedy and Gentle (1980) proposed an acceptance-rejection method to sample from an arbitrary density $h(z)$. One must find a positive, integrable function $d(z)$ that dominates $h(z)$, i.e., $0 \leq h(z) \leq d(z)$ for any $z$. In this case, $d(z)$ is called an upper envelope for $h(z)$ or a majorizing function. Derive a density $g(z)$ from $d(z)$ by normalizing it, i.e.,

$$g(z) = \frac{d(z)}{\int d(s)ds}.$$
Using $d(z)$ and $g(z)$, a sample $z$ from $h(z)$ is generated as follows.

1. Generate the pair $(u, v)$ independently, $u \sim \text{Uniform}(0, 1)$, $v \sim g(v)$.

2. If $u \leq h(v)/d(v)$ then accept $z = v$; otherwise, go to 1 and repeat until one sample $z$ is obtained.

2.3.2 Sampling from a SNP Density

Gallant and Tauchen (1992) and Chen et al. (2002) described how to generate a random sample from a SNP density with the normal “kernel” and fixed degree of truncation $K$. We do not reproduce the sampling scheme here, but refer readers to these papers for details. Here, we similarly introduce how to generate a random sample from a SNP density with the exponential “kernel.” We apply the algorithm proposed by Gallant and Tauchen (1992). Our objective density is the SNP density with exponential “kernel” and fixed degree of $K$ given by

$$h_K(z) = P_K^2(z; a)e^{-z}, \quad z > 0,$$

where

$$P_K(z; a) = \sum_{|\lambda| = 0}^{K} a_{\lambda} z^{\lambda}.$$

Since

$$\sum_{|\lambda| = 0}^{K} a_{\lambda} z^{\lambda} \leq \sum_{|\lambda| = 0}^{K} |a_{\lambda}| |z|^{|\lambda|},$$

we can choose the upper envelope function as

$$d_K(z) = \left\{ \sum_{|\lambda| = 0}^{K} |a_{\lambda}| |z|^{|\lambda|} \right\}^2 e^{-z}.$$
We can easily obtain the density $g_K(z)$, which is a mixture of Gamma distribution and is given by

$$g_K(z) = \frac{d_K(z)}{\int d_K(s)ds} = \sum_{|\lambda|=0}^{K} \sum_{|\gamma|=0}^{K} \omega_{\lambda\gamma} \text{Gamma}(z; \lambda + \gamma + 1, 1),$$

where $\text{Gamma}(z; \lambda + \gamma + 1, 1)$ is denoted as Gamma density with shape parameter $\lambda + \gamma + 1$ and scale parameter 1, and

$$\omega_{\lambda\gamma} = \frac{|a_\lambda||a_\gamma|\Gamma(\lambda + \gamma + 1)}{\sum_{|\lambda|=0}^{K} \sum_{|\gamma|=0}^{K} |a_\lambda||a_\gamma|\Gamma(\lambda + \gamma + 1)}.$$

In order to sample from $g_K(z)$, we need to obtain a sample $(\lambda, \gamma)$ from the set $F(\lambda, \gamma) = \{(\lambda, \gamma) : 0 \leq |\lambda| \leq K, 0 \leq |\gamma| \leq K\}$. This can be done as follows. Let the elements of $F(\lambda, \gamma)$ be ordered in an arbitrary way so that they can be indexed by the sequence $(\lambda, \gamma)_j$, where $j = 0, 1, \ldots, J$. Let $\omega_j = \omega_{\lambda\gamma}$ where $(\lambda, \gamma) = (\lambda, \gamma)_j$. Generate $u \sim \text{Uniform}(0, 1)$. Find the largest $L$ such that $\sum_{j=0}^{L} \omega_j \leq u$. Then let $(\lambda, \gamma) = (\lambda, \gamma)_L$. For this given $(\lambda, \gamma)$, generate $z$ from $\text{Gamma}(z; \lambda + \gamma + 1, 1)$. Then $z$ is a sample point from $g_K(z)$. The modified rejection sampling algorithm to generate a sample from SNP density $h_K(z)$ is as follows.

1. Sample $(\lambda, \gamma)$ from the set $F(\lambda, \gamma)$.
2. For the given $(\lambda, \gamma)$, sample $z$ from $g_K(z)$.
3. Generate $u$ from $\text{Uniform}(0, 1)$ independently. If $u \leq h_K(z)/d_K(z)$, then accept $z$; otherwise, go to 1 and repeat until one sample point is obtained.

We use two examples to illustrate the performance of the proposed acceptance-rejection sampling method. For $q = 1$ and $K = 1$, let $h_1(z) = P_1^2(z; a)e^{-z}$, where
\( P_1(z; a) = a_1 + a_2 z, \) with \( \psi = (-0.7), \) we obtain \( a = (-1.409, 0.765)^T. \) We generate 30000 samples from this SNP density, and the histogram of a random samples is shown in Figure 2.1. We can see the shape of the histogram matches perfectly to the truth, which shows the algorithm works well and the random samples generated using the above algorithm come from the true SNP density. However, the acceptance rate is about 19\%, which is low. In order to increase the acceptance rate, we need to improve the upper envelope function \( d_K(z). \) We leave this as a topic for future research; for our purposes in this dissertation, the acceptance rate is adequate.
As another example, take $q = 1$ and $K = 2$, $h_2(z) = P_2^q(z; a)e^{-z}$, where $P_2(z; a) = a_1 + a_2 z + a_3 z^2$. Letting $\psi = (-0.7, -1.2)^T$, we have $a = (0.346, -1.267, 0.139)^T$. Figure 2.2 shows the true density and a histogram of 30000 samples based on the method we proposed above. The shape of the histogram again matches perfectly to the truth.
Figure 2.2: Histogram of 30000 random samples generated from a SNP density and true density ($K = 2, q = 1$)
Chapter 3

Accelerated Failure Time
Regression Model for Clustered Correlated Data

In this chapter, we describe our proposed AFT model for clustered correlated survival data based on the SNP representation, and propose a likelihood-based approach to inference implemented via an EM algorithm.

3.1 Accelerated Failure Time Regression Model

Let $T_{ij}$ be the $j$th failure time in cluster $i$, $i = 1, \ldots, n$, $j = 1, \ldots, m_i$. $T_{ij}$ might be interval censored, so is known to only lie in an interval $[L_{ij}, R_{ij}]$; left censored at $R_{ij}$ (which is a special case of interval censoring with $L_{ij} = 0$); right censored at $L_{ij}$
(set $R_{ij} = \infty$); or observed (set $T_{ij} = L_{ij} = R_{ij}$). Set $\Delta_{ij} = 1$ if the failure time is observed, otherwise set $\Delta_{ij} = 0$. The observed i.i.d. data are $(L_i, R_i, \Delta_i), i = 1, \ldots, n,$ where $L_i = (L_{i1}, \ldots, L_{im_i})^T, R_i = (R_{i1}, \ldots, R_{im_i})^T,$ and $\Delta_i = (\Delta_{i1}, \ldots, \Delta_{im_i})^T$; let $L = (L_1, \ldots, L_n)^T, R = (R_1, \ldots, R_n)^T$ and $\Delta = (\Delta_1, \ldots, \Delta_n)^T$; and let $X_{ij}$ be covariates. In this chapter, we assume $T_{ij}$ represents the same type of time-to-event for all $j$ (“Parallel” events). Although variables are independent across $i$, for given $i$, $T_{ij}$ may be correlated across $j$.

The objective of an analysis is to characterize the relationship between $X_{ij}$ and $T_{ij}$ on the basis of the observed data within the framework of a regression model that takes appropriate account of within-cluster dependence. A natural way to address dependence is through specification of a transformed failure time model

$$g(T_{ij}) = m(X_{ij}, \beta, b_i) + e_{ij}$$

(3.1)

where $g(\cdot)$ is a monotone transformation from $(0, \infty)$ to the real line; $m(X_{ij}, \beta, b_i)$ depends on $\beta(p \times 1), X_{ij},$ and $b_i$; the $e_{ij}$’s are i.i.d. within-cluster “error” terms assumed to have distribution with a density satisfying mild “smoothness” conditions, with density, distribution, and survival functions $f_0$, $F_0$, and $S_0$; and i.i.d. scalar random effects $b_i$ are normally distributed with mean zero and variance-covariance matrix. The AFT model (1.2) in Section 1.2 is a special case of (3.1) with $g(t) = \log(t)$ and $m(X_{ij}, \beta, b_i) = X_{ij}^T \beta + S_{ij}^T b_i$, where $S_{ij}$ is an additional design vector for random effect $b_i$. Depending on whether $m(X_{ij}, \beta, b_i)$ is linear in $b_i$, the interpretation of $\beta$ as a subject-specific or population-averaged parameter is as for usual mixed models.

The methods we propose in this dissertation may be adapted to general models as
in (3.1). For definiteness, we develop a method in the particular case of the cluster-specific AFT model with random intercept given by

$$\log(T_{ij}) = X_{ij}^T \beta + b_i + e_{ij},$$

(3.2)

where $b_i$ is a scalar, $N(0, \sigma_b^2)$ random effect. In our approach, the density $f_0$ (hence $S_0$) is approximated by the SNP representation. Because no intercept is included in the proposed subject-specific AFT model, it is not necessary to assume the within-cluster error has mean zero.

### 3.2 SNP Representation of the Error Density

Based on the review of SNP representation in Chapter 2, for a location parameter $\mu$ and a scale parameter $\gamma$, and a normal “kernel” density for the SNP suggests the representation of within-cluster error term $e_{ij}$ as

$$\mu + \gamma Z_{ij},$$

(3.3)

where $Z_{ij}$ is assumed to have a density $h(z) \in \mathcal{H}$, so that $f_0 \in \mathcal{H}$. Under these assumptions, for fixed $K$, we might approximate $h(z)$ by the “standard” SNP density representation with normal “kernel” $h_K(z) = P_K^2(z)\xi_1(z)$, where $\xi_1(z) = \varphi(z; 0, 1)$ denotes the $N(0, 1)$ density. When $K = 0$, $h_K(z)$ is the standard normal $N(0, 1)$ density. Thus, if the error term truly comes from normal distribution with mean $\mu$ and variance $\gamma^2$, the SNP representation with normal “kernel” and $K = 0$ as in (3.3) should fit the data perfectly.

For some true densities, particularly those that are very skewed, in order to achieve
a satisfactory approximation, we need a larger $K$. Accordingly, the potential to incorporate “kernel” other than normal suggests an alternative formulation of model (3.2). Consider the model

$$T_{ij} = \exp(X_{ij}^T \beta + b_i)e^{e_{ij}} = \tilde{m}(X_{ij}, \beta, b_i)\delta_{ij},$$

(3.4)

where $\tilde{m}(X_{ij}, \beta, b_i)$ takes postive value; and $\delta_{ij} = e^{e_{ij}}$ has density, distribution and survival functions $\tilde{f}_0, \tilde{F}_0,$ and $\tilde{S}_0$. $\delta_{ij}$ can be further written as $\mu^*Z_{ij}^\gamma$, where $Z_{ij}$ has density $h(z) \in H$ with support on $(0, \infty)$. We may represent $h(z)$ by a truncated SNP density with some “kernel” having moment generating function and support on $(0, \infty)$. A natural choice is the standard exponential distribution, so that $h_K(z) = P^2_K(z)e^{-z}$, where we also require $\int_0^\infty h_K(z)dz = 1$. Note that, in effect, this representation of $h$ is a mixture of gammas; the richness of the mixture is controlled by $K$. Therefore, we can represent the within-cluster error term $e_{ij}$ as

$$\mu + e^{\gamma \log Z_{ij}},$$

(3.5)

where $Z_{ij}$ has density $h(z)$. For fixed $K$, we approximate $h(z)$ by the “standard” SNP density representation with exponential “kernel” $h_K(z) = P^2_K(z)\xi_1(z)$, where $\xi_1(z) = e^{-z}$. In Section 3.8, we show that when $\delta_{ij}$ comes from exponential distribution with some scale parameter or a Weibull distribution with some shape parameter $\alpha$ and scale parameter $\nu$, the SNP representation with exponential “kernel” and $K = 0$ should fit perfectly.

Note: If we represent $e_{ij} = \mu + \gamma \log Z_{ij}$ for the exponential “kernel,” $\gamma$ should be positive. However, it has been our experience that, unlike for (3.3), $\gamma$ sometimes can be estimated by some negative value. In order to avoid this numerical problem, we
substitute $\gamma$ by $e^\gamma$ in (3.5). This substitution does not affect the proposed model.

Let $\theta = (\mu, \gamma, \psi^T, \beta^T)^T$, then $\zeta = (\theta^T, \sigma_b^2)^T$ be the $(K + p + 3)$ dimensional vector of parameters characterizing the model (3.2) and (3.3) or the model (3.2) and (3.5). We denote the conditional density and survival function of $T_{ij}$ given the random effect $b_i$ by $f(t|b_i; \theta)$ and $S(t|b_i; \theta)$, respectively.

### 3.3 Likelihood Function

We now consider likelihood-based inference on the regression parameter, random effect variance component, and SNP parameters describing the error density for fixed $K$ and “kernel.” In this dissertation, the likelihood is conditioned on the covariates $X_{ij}$, and we suppress conditioning on covariates for simplicity. For fixed $K$ and “kernel,” under the models given in (3.2), (3.3) and (3.5), the contribution to the likelihood for $\zeta$ for cluster $i$ based on the data $(L_i, R_i, \Delta_i)$ is, in obvious notation,

$$
\mathcal{L}(\zeta; L_i, R_i, \Delta_i) = \int f(L_i, R_i, \Delta_i; b_i; \zeta) db_i = \int f(L_i, R_i, \Delta_i; b_i; \theta) \varphi(b_i; 0, \sigma_b^2) db_i,
$$

(3.6)

where $\varphi(b_i; 0, \sigma_b^2)$ is the normal density with mean 0 and variance $\sigma_b^2$, and

$$
\log\{f(L_i, R_i, \Delta_i; b_i; \theta)\} = \sum_{j=1}^{m_i} [\Delta_{ij} \log\{f(L_{ij}|b_i; \theta)\} + (1 - \Delta_{ij}) \log\{S(L_{ij}|b_i; \theta) - S(R_{ij}|b_i; \theta)\}].
$$

Here, for the normal “kernel” SNP representation,

$$
f(L_{ij}|b_i; \theta) = \frac{1}{\gamma L_{ij}} P_K^2(r_{ij}) \varphi(r_{ij}; 0, 1),
$$
\[ S(L_{ij}|b_i; \theta) = \int_{r_{ij}}^{\infty} P_K^2(z) \varphi(z; 0, 1) \, dz, \]

\[ r_{ij} = \frac{\log(L_{ij}) - X^T_{ij} \beta - b_i - \mu}{\gamma}, \]

\[ S(R_{ij}|b_i; \theta) = \int_{r_{ij}^*}^{\infty} P_K^2(z) \varphi(z; 0, 1) \, dz, \]

\[ r_{ij}^* = \frac{\log(R_{ij}) - X^T_{ij} \beta - b_i - \mu}{\gamma}; \]

and

\[ f(L_{ij}|b_i; \theta) = \frac{1}{L_{ij}e^\gamma} P_K^2(s_{ij}) e^{-s_{ij}} s_{ij}, \]

\[ S(L_{ij}|b_i; \theta) = \int_{s_{ij}}^{\infty} P_K^2(z) e^{-z} \, dz, \]

\[ s_{ij} = \exp \left\{ \frac{\log(L_{ij}) - X^T_{ij} \beta - b_i - \mu}{e^\gamma} \right\}, \]

\[ S(R_{ij}|b_i; \theta) = \int_{s_{ij}^*}^{\infty} P_K^2(z) e^{-z} \, dz, \]

\[ s_{ij}^* = \exp \left\{ \frac{\log(R_{ij}) - X^T_{ij} \beta - b_i - \mu}{e^\gamma} \right\} \]

for the exponential “kernel.”

The conditional survival function, \( S(L_{ij}|b_i; \theta) \), may be written as a linear combination of integrals of the form \( I(k, r_{ij}) = \int_{r_{ij}}^{\infty} z^k \xi_1(z) \, dz \), and similarly for \( S(R_{ij}|b_i; \theta) \), where \( \xi_1(z) \) is either standard normal density or standard exponential density. For the normal “kernel,” clearly, \( I(0, r) = 1 - \Phi(r) \), where \( \Phi(\cdot) \) is the standard normal cdf, and \( I(1, r) = \varphi(r; 0, 1) \). For \( k \geq 2 \), it is straightforward to show that these integrals may be calculated recursively using \( I(k, r) = r^{k-1} \varphi(r; 0, 1) + (k-1)I(k-2, r) \).

For the exponential “kernel,” clearly, \( I(0, r) = 1 - \Phi_e(r) \), where \( \Phi_e \) is the standard exponential cdf. For \( k \geq 1 \), it may be shown that these integrals may be calculated recursively as \( I(k, r) = r^k e^{-r} + kI(k-1, r) \).
The loglikelihood for $\zeta$ may be written as

$$
\ell(\zeta; L, R, \Delta) = \sum_{i=1}^{n} \log \{ \mathcal{L}(\zeta; L_i, R_i, \Delta_i) \}, \quad (3.7)
$$

where $\mathcal{L}(\zeta; L_i, R_i, \Delta_i)$ is defined as before.

Inference may be based on maximizing the loglikelihood (3.7). However, (3.7) cannot be evaluated analytically in general. We propose to maximize (3.7) in $\zeta$ via a Monte Carlo EM (MCEM) algorithm similar to that of Booth and Hobert (1999) using rejection sampling.

### 3.4 Implementations

#### 3.4.1 Monte Carlo EM Algorithm

The challenge in maximizing (3.7) is that the integral in (3.6) can not be evaluated in a closed form. One approach that has proved successful in such situations is to use the EM algorithm treating the random effects $b_i$ as “missing” data and $(L, R, \Delta)$ as “observed” data (e.g., Searle, Casella, and McCulloch, 1992, Chapter 8). Then the “complete” data $(L, R, \Delta, b)$ have the joint density function $f(L, R, \Delta, b; \zeta)$. We now outline how the algorithm proceeds for our problem.

Given the $r$th iterate estimate $\zeta^{(r)}$, at the $(r + 1)$th iteration, the E–step involves the calculation of

$$
Q(\zeta|\zeta^{(r)}) = E \{ \log f(L, R, \Delta, b; \zeta) | L, R, \Delta; \zeta^{(r)} \} = Q_1(\zeta|\zeta^{(r)}) + Q_2(\zeta|\zeta^{(r)}), \quad (3.8)
$$
where

\[ Q_1(\zeta|\zeta^{(r)}) = \sum_{i=1}^{n} \sum_{j=1}^{m_i} E[\Delta_{ij} \log \{f(L_{ij}|b_i;\theta)\} \right \} \]

\[ + (1 - \Delta_{ij}) \log \{S(L_{ij}|b_i;\theta) - S(R_{ij}|b_i;\theta)\}|L_{ij}, R_{ij}, \Delta_{ij}, \zeta^{(r)}], \]

and

\[ Q_2(\zeta|\zeta^{(r)}) = E \left( -\frac{n}{2} \log 2\pi - n \log \sigma_b - \sum_{i=1}^{n} \frac{b_i^2}{2\sigma^2_b} \right) \]

The M–step consists of maximizing \( Q(\zeta|\zeta^{(r)}) \) with respect to \( \zeta \) to yield the new update \( \zeta^{(r+1)} \). If we set up a starting value \( \zeta^{(0)} \), we iterate the process to maximize \( Q(\zeta|\zeta^{(r)}) \) to convergence; under the regularity conditions, the value at convergence maximizes the likelihood function (3.7).

Usually, there is no closed form for expression for (3.8), as it requires knowledge of the conditional distribution of \( b \) given \( (L, R, \Delta) \) evaluated at \( \zeta^{(r)} \), which in turn requires knowledge of the marginal likelihood. In order to circumvent this difficulty, much effort (Wei and Tanner, 1990; McCulloch, 1997; Booth and Hobert, 1999) has been made to use a Monte Carlo approximation to the required integration at the E–step. Specifically, if it is possible to obtain a random sample \( (b^{(1)}, b^{(2)}, \ldots, b^{(L)})^T \) from \( f(b|L, R, \Delta; \zeta^{(r)}) \), then (3.8) may be approximated by

\[ Q_L(\zeta|\zeta^{(r)}) \]

\[ = L^{-1} \sum_{i=1}^{L} \left\{ \sum_{i=1}^{n} \sum_{j=1}^{m_i} \left[ \Delta_{ij} \log \{f(L_{ij}|b_i^{(l)};\theta)\} \right] \right. \]

\[ + (1 - \Delta_{ij}) \log \{S(L_{ij}|b_i^{(l)};\theta) - S(R_{ij}|b_i^{(l)};\theta)\} \left. \right] - \left\{ n \log \sigma_b^2 + \sigma_b^{-2} \sum_{i=1}^{n} (b_i^{(l)})^2/2 \right\} \]

36
yielding a so-called Monte Carlo EM (MCEM) algorithm. Thus

\[ \frac{\partial Q_L}{\partial \sigma_b^2} = \frac{1}{L} \sum_{l=1}^{L} \left( \frac{n}{\sigma_b^2} - \frac{1}{\sigma_b^4} \sum_{i=1}^{n} b_{i(l)}^2 \right) = 0, \]

which yields

\[ \sigma_b^{2(r+1)} = \frac{\sum_{l=1}^{L} \sum_{i=1}^{n} b_{i(l)}^2}{nL}. \]

Letting

\[ Q_{L_1} = L^{-1} \sum_{l=1}^{L} \left\{ \sum_{i=1}^{n} \sum_{j=1}^{m_i} \left[ \Delta_{ij} \log \{ f(L_{ij}|b_{i(l)}; \theta) \} 
+ (1 - \Delta_{ij}) \log \{ S(L_{ij}|b_{i(l)}; \theta) - S(R_{ij}|b_{i(l)}; \theta) \} \right] \right\}. \]

\( \theta^{(r+1)} \) may be obtained by maximizing \( Q_{L_1} \). In our implementations, we have used Intel Fortran 95 IMSL library DUMPOL subroutine to maximize \( Q_{L_1} \) under the Microsoft Visual Studio.net integrated development environment (IDE). By independence, to obtain a sample from \( f(b|L,R,\Delta; \zeta^{(r)}) \), one may sample from the conditional distribution of \( b_i \) given \( L_i, R_i, \Delta_i \) evaluated at \( \zeta^{(r)}, f(b_i|L_i,R_i,\Delta_i; \zeta^{(r)}) \), say, for each \( i \).

Incorporating the Monte Carlo approximation into the EM algorithm gives the full MCEM algorithm as follows.

1. Choose starting values \( \zeta^{(0)} \). Set \( r = 0 \).

2. At iteration \( (r + 1) \), generate a random sample \( b^{(l)} \) from \( f(b|L,R,\Delta; \zeta^{(r)}) \), \( l = 1, \ldots, L \).

3. Obtain \( \zeta^{(r+1)} \) by maximizing \( Q_L(\zeta|\zeta^{(r)}) \).
4. If convergence is achieved, set $\zeta^{(r+1)}$ to be the maximum likelihood estimate $\hat{\zeta}$; otherwise, set $r = r + 1$ and return to 2.

To complete the MCEM algorithm, we need to know how to generate such random samples from the conditional distribution $f(b_i|L_i, R_i, \Delta_i; \zeta^{(r)})$. McCulloch (1997) suggested the Metropolis-Hastings algorithm and Booth and Hobert (1999) proposed to use a rejection sampling algorithm. In our paper, we adopt the rejection sampling scheme (Booth and Hobert, 1999).

### 3.4.2 Rejection Sampling

For our problem, let $\tau_i = \sup_{b_i} f(L_i, R_i, \Delta_i; b_i; \theta^{(r)})$. Then a random sample $b_i$ may be obtained by the following steps.

1. Generate a random sample $b^*_{i}$ from $N(0, \sigma_b^{2(r)})$.

2. Sample $u \sim \text{Uniform}(0, 1)$ independently. If $u \leq f(L_i, R_i, \Delta_i; b^*_{i}; \zeta^{(r)})/\tau_i$, accept $b_i = b^*_{i}$; otherwise, return 1 and repeat until a sample $b_i$ is obtained.
It is straightforward to see that \( b_i \) generated by the above rejection sampling scheme is indeed a random sample from \( f(b_i | L_i, R_i, \Delta_i; \zeta^{(r)}) \) by noting that

\[
\Pr(b_i \leq t) = \frac{\Pr\{(b_i^*, u) : b_i^* \leq t, u \leq f(L_i, R_i, \Delta_i|b_i^*; \zeta^{(r)})/\tau_i\}}{\Pr\{(b_i^*, u) : u \leq f(L_i, R_i, \Delta_i|b_i^*; \zeta^{(r)})/\tau_i\}}
\]

\[
= \frac{\int_{b_i^* \leq t} \int_0^{f(L_i, R_i, \Delta_i|b_i^*; \zeta^{(r)})/\tau_i} f(b_i^*; \delta^{(r)}) \, db_i^* \, du}{\int_{b_i^* \leq \infty} \int_0^{f(L_i, R_i, \Delta_i|b_i^*; \zeta^{(r)})/\tau_i} f(b_i^*; \delta^{(r)}) \, db_i^* \, du}
\]

\[
= \frac{\int_{b_i^* \leq t} f(L_i, R_i, \Delta_i|b_i^*; \zeta^{(r)}) \, f(b_i^*; \delta^{(r)}) \, db_i^*}{\int_{b_i^* \leq \infty} f(L_i, R_i, \Delta_i|b_i^*; \zeta^{(r)}) \, f(b_i^*; \delta^{(r)}) \, db_i^*}
\]

\[
= \frac{\int_{b_i^* \leq t} f(L_i, R_i, \Delta_i; b_i^*; \zeta^{(r)}) \, db_i^*}{f(L_i, R_i, \Delta_i; \zeta^{(r)})}
\]

\[
= \int_{b_i^* \leq t} f(b_i^*|L_i, R_i, \Delta_i; \zeta^{(r)}) \, db_i^*.
\]

### 3.5 Choosing \( K \) and the “Kernel” Density

The degree of the SNP polynomial, \( K \), acts as a tuning parameter similar to the bandwidth in kernel density estimation to control the degree of smoothness. A larger \( K \) model gives more flexibility for representing the error term distribution, but choosing a larger \( K \) will result in an inefficient representation. Thus, it is desirable to choose \( K \) according to some criterion that balances the suitability of the fit and the number of parameters. Davidian and Gallant (1993), Zhang and Davidian (2001), Chen et al. (2002), Zhang and Davidian (2007), and Doehler and Davidian (2008) advocated inspection of information criteria of the form \(- \log L(\hat{\zeta}; L, R, \Delta) + c_N p_{net}\). In our problem, \( \hat{\zeta} \) is the MCEM estimate, \( p_{net} \) is the dimension of the parameters to be estimated, \( c_N \equiv 1 \) yields the Akaike’s information criterion (AIC), \( c_N = \frac{1}{2} \log N \) gives Schwarz’s
Bayesian information criterion (BIC), and $c_N = \log \log N$ gives the Hannan-Quinn criterion (HQ). For all of AIC, BIC, and HQ, smaller values are preferred. AIC tends to yield larger models and BIC smaller models, with HQ intermediate. Here we use HQ as our information criterion to choose $K$ and the “kernel.” Because the data are censored, use of the total number of observations may be suspect. Koo et al. (1999) and Volinsky and Raftery (2000) suggested using instead the number of uncensored events instead of number of observations in the AIC, BIC or HQ criterion. We have investigated the use of different such choices and found that both the number of the uncensored events and number of the observations yield very similar results in choosing $K$ and the “kernel” density. Thus, for simplicity, the results in this dissertation are all based on using the number of observations in the HQ criterion. Our preliminary results also show that $K = 3$ is seldom chosen and $K = 2$ is sufficient to represent adequately departures from normality and exponentiality. Thus we advocate fitting the model using $K = 0, 1, 2$ for both the normal “kernel” and exponential “kernel” and selecting the “best” combination minimizing the HQ information criterion. To compute HQ, $\log L(\hat{\zeta}; L, R, \Delta)$ may be approximated by Gauss-Hermite quadrature with 40 points.

3.6 Standard Errors and Confidence Intervals

After convergence is achieved, the maximum likelihood estimates (MLEs) of $\zeta$ are obtained. The value of loglikelihood at the final estimates may be calculated by using Gauss-Hermite quadrature; we have used 40 quadrature points. The gradient
and observed Fisher information may be then obtained by numerical differentiation; we have used Fortran double precision subroutine DCDGRD. We treat $K$ and “kernel” base density as predetermined following Gallant and Tauchen (1992) and Zhang and Davidian (2007). The asymptotic variance-covariance matrix may be approximated by the inverse of the observed Fisher information matrix acting as if the chosen $\ell(\zeta; L, R, \Delta)$ were the loglikelihood under a predetermined parametric model. The square root of each diagonal element will be the standard error for the corresponding component of the MLE of $\zeta$. Although we choose $K$ and “kernel” density adaptively, which seems to invalidate this practice, simulation results and examples shown in Chapter 4 demonstrate that this approach yields reliable inferences in realistic sample sizes.

The Supplementary Materials of Zhang and Davidian (2007), web Appendix A, offer a detailed discussion of this method of obtaining standard errors and what is known about its performance. There is, however, currently no formal theoretical justification for treating $K$ and the “kernel” density as fixed. Wald confidence intervals for the true values of elements $\zeta$ can be constructed based on this approach. The delta method may be used to calculate the standard error of the expectation and variance of the error term $e_{ij}$.
3.7 Hypothesis Testing

3.7.1 Hypothesis Testing for Regression Parameters

It will often be of interest to test hypotheses regarding elements of the regression parameter vector $\beta$. For example, one element of $\beta$ may correspond to a treatment effect (perhaps after adjusting for the effects of other covariates). We advocate Wald inference for this purpose; specifically, one may construct a test statistic as the estimated parameter divided by the estimated standard error calculated as described in Section 3.6.

3.7.2 Hypothesis Testing for Variance Components

For clustered survival data, the failure times within each cluster might be correlated, which is taken into account in the cluster-specific AFT model by a random intercept, $b_i$, assumed $N(0, \sigma^2_b)$. It is often of interest to test whether the potential correlation among the survival times in each cluster is non-zero, which suggests that a model treating these as mutually independent would be inappropriate. The correlation will be 0 if $\sigma^2_b = 0$. Classical inference based on the likelihood of the marginal model implied by (3.2) allows the variance of the random effect to be negative. However, when the link between our model and marginal model is preserved, it makes sense to restrict $\sigma^2_b$ to have a nonnegative value (Verbeke and Molenberghs, 2003). In this situation, one is interested in a one-sided test of the null hypothesis under this
constrained case, i.e.,

\[ H_0 : \sigma_b^2 = 0 \ vs. \ H_0 : \sigma_b^2 > 0. \] (3.9)

The null hypothesis is now on the boundary of the parameter space and classical inference no longer holds. An appropriate test statistic for (3.9) is required. This one-sided approach has received considerable attention in the case of the likelihood ratio test (Self and Liang, 1987; Stram and Lee, 1994; Verbeke and Molenberghs, 2003). Verbeke and Molenberghs (2003) considered this hypothesis for the linear mixed model and proved that, under \( H_0 \), the asymptotic distribution of the one-sided likelihood ratio test statistic is an equal (probability 0.5) mixture of two chi-squared distributions with degrees of freedom 0 and 1.

Given the similarity of our model to this model, we adopt the likelihood ratio test for testing (3.9). To implement the test, we fit the “full” model with \( \sigma_b^2 \) estimated, determine \( K \) and the “kernel.” Then the “reduced” model is fitted with \( \sigma_b^2 = 0 \) using same \( K \) and “kernel.” In the “reduced” model, the density of the error term is still estimated by SNP density representation. Let \( \Theta_0 \) be the sub-parameter space with \( \sigma_b^2 = 0 \) of the parameter space \( \Theta \), then the likelihood ratio test statistic is defined as

\[
LRT = -2 \log \Lambda(L, R, \Delta),
\]

where

\[
\Lambda(L, R, \Delta) = \frac{\sup \{ L(\zeta; L, R, \Delta) : \zeta \in \Theta_0 \}}{\sup \{ L(\zeta; L, R, \Delta) : \zeta \in \Theta \}}.
\]

Under the null hypothesis, we can compare the likelihood ratio test statistic with the level-\( \alpha \) critical point of a mixture of \( \chi^2(0) \) and \( \chi^2(1) \) with equal probability 0.5 to determine if we should reject the null hypothesis or not.

We reject \( H_0 \) if \( LRT \geq C \). We derive the level-\( \alpha \) critical point \( C \). Let \( \tilde{Z}_1 \) be a
standard normal \( N(0, 1) \) random variable. Then, under the null hypothesis, likelihood ratio test statistic can be written as

\[
\tilde{Z}_1^2 I(\tilde{Z}_1 \geq 0),
\]

so that,

\[
P\{\tilde{Z}_1^2 I(\tilde{Z}_1 \geq 0) \geq C\} = \alpha,
\]

which yields

\[
\frac{1}{2} P(\tilde{Z}_1^2 \geq C) = \alpha.
\]

Therefore,

\[
P(\tilde{Z}_1^2 \geq C) = 2\alpha.
\]

Then we will reject \( H_0 \) if \( LRT \geq \chi^2_{1,2\alpha} \).

In Chapter 4, we present simulations showing the performance of this test.

### 3.8 Technical Details

We show that the SNP representation with exponential “kernel” and \( K = 0 \) should perfectly fit when \( \delta_{ij} \) comes from Weibull distribution or exponential distribution. Because the exponential distribution is a special case of the Weibull distribution, here we only prove the case when \( \delta_{ij} \) comes from a Weibull distribution.

Let \( \delta_{ij} \sim \text{Weibull}(\alpha, \nu) \). Then the density of \( \delta_{ij} \) is

\[
g(\delta) = \alpha \nu^{-\alpha} \delta^{\alpha-1} e^{-(\frac{\delta}{\nu})^\alpha}.
\]  

(3.10)
Let $Y = \log \delta_{ij}$, so that $\delta_{ij} = e^Y$. Then the density of $Y$, $f(y)$, is

$$f(y) = g(e^y)e^y$$

$$= \alpha \nu^{-\alpha} (e^y)^{\alpha-1} e^{-(\frac{e^y}{\nu})^\alpha} e^y$$

$$= \alpha \nu^{-\alpha} e^{\alpha y - \nu^{-\alpha} e^{\alpha y}},$$

i.e.,

$$f(y) = \alpha \nu^{-\alpha} e^{\alpha y - \nu^{-\alpha} e^{\alpha y}}. \quad (3.11)$$

If we represent $Y = \log \delta_{ij}$ by a SNP density with exponential “kernel,” then

$$Y = \mu + e^\gamma \log Z,$$

where $Z$ has SNP density $P^2_K(z)e^{-z}$. Therefore,

$$P(Y \leq y) = P(\mu + e^\gamma \log Z \leq y)$$

$$= P(\log Z \leq \frac{y - \mu}{e^\gamma})$$

$$= P(Z \leq e^{\frac{y - \mu}{e^\gamma}}).$$

Then

$$f(y) = P^2_K(e^{\frac{y - \mu}{e^\gamma}}) \exp\{-e^{\frac{y - \mu}{e^\gamma}}\} e^{\frac{y - \mu}{e^\gamma}} \frac{1}{e^\gamma}$$

$$= P^2_K(e^{\frac{y - \mu}{e^\gamma}}) \exp\{\frac{y}{e^\gamma} - e^{\frac{y - \mu}{e^\gamma}}\} \frac{1}{e^\gamma e^\gamma}. $$

If $K = 0$, then

$$f(y) = \exp\{\frac{y}{e^\gamma} - e^{\frac{y}{e^\gamma}}\} \frac{1}{e^\gamma e^\gamma}. \quad (3.12)$$

If we compare two representations of the density of $Y$, $f(y)$, (3.11) and (3.12), and let $\alpha = \frac{1}{e^\gamma}$ and $\nu = e^\mu$, then it is clear that the two densities are exactly same. Therefore we can conclude that, when $\delta_{ij}$ comes from a Weibull distribution, the SNP with exponential “kernel” and $K = 0$ should perfectly fit.
Chapter 4

Simulations and Examples for Clustered Correlated Data

We conducted simulation studies with right- and interval-censored data to investigate the performance of the methods proposed. In Section 4.1, we show the simulation results for maximum likelihood estimation of the model parameters. In Section 4.2, we address the performance of the proposed likelihood ratio test for hypothesis testing for variance components, given in Section 3.7.2. In our proposed cluster-specific AFT model, we assume that the random effect $b_i$ comes from normal distribution with mean zero and variance $\sigma^2_b$. However, if this assumption is violated, it is important to evaluate the effect on inference on model parameters. A full treatment of this issue is beyond the scope of this dissertation, and is a topic for future research; however, we present a simulation addressing this in Section 4.3.
4.1 Maximum Likelihood Estimation

4.1.1 Designs of Simulations

We simulated several scenarios where the data are subject right- or interval-censoring. Under each scenario, we generated 100 Monte Carlo data sets, each with 100 clusters and 6 observations in each cluster. We generated $T_{ij}, i = 1, \ldots, n = 100, j = 1, \ldots, m_i = 6$ from model (3.2) described in Chapter 3,

$$\log(T_{ij}) = \beta X_{ij} + b_i + e_{ij} = \beta X_{ij} + b_i + \log(\delta_{ij}).$$

(4.1)

As shown in Section 3.2, $\delta_{ij}$ is equal to $e^{e_{ij}}$, and we write the error term either as $e_{ij}$ or $\log \delta_{ij}$ as appropriate. The true value of $\beta$ was always 2.0 for right-censored data, and $b_i$ is drawn from $N(0, \sigma_b^2)$. In all of the simulations, the value of $\sigma_b^2$ was chosen so that the within-cluster correlation coefficient within cluster is 0.50, where the correlation coefficient is defined as $\sigma_b^2/($$\sigma_b^2 + \sigma_e^2$), and $\sigma_e^2 = \text{Var}(e_{ij})$. For right-censored data, the censoring time $L_{ij}$ was independently drawn from a Uniform distribution on $(0, u)$, where $u$ is chosen so that the censoring rate is always about 30%. The covariate $X_{ij}$ was drawn either from a Uniform(0, 1) distribution or Bernoulli(0.5) distribution. When $X_{ij}$ was drawn from Uniform distribution, it may be different for each individual in a cluster. But when $X_{ij}$ was drawn from Bernoulli distribution, it may be same or different within a cluster as described below.

We carried out simulations under the following scenarios:

1. $e_{ij} \sim N(\mu, \sigma_e^2)$ with $\mu = -1.5, \sigma_e^2 = 2.0, \sigma_b^2 = 2.0; X_{ij} \sim \text{Uniform}(0, 1); \text{ and } L_{ij} \sim \text{Uniform}(0, 5)$. 

47
2. $e_{ij} \sim N(\mu, \sigma^2_e)$ with $\mu = -1.5$, $\sigma^2_e = 2.0$, $\sigma^2_b = 2.0$; $X_{ij} \sim \text{Bernoulli}(0.5)$, where the covariate may be different for each individual in a cluster; and $L_{ij} \sim \text{Uniform}(0, 6)$.

3. $e_{ij} \sim N(\mu, \sigma^2_e)$ with $\mu = -1.5$, $\sigma^2_e = 2.0$, $\sigma^2_b = 2.0$; $X_{ij} \sim \text{Bernoulli}(0.5)$, where the covariate is same for each individual in a cluster; and $L_{ij} \sim \text{Uniform}(0, 6)$.

4. $e_{ij}$ were drawn from a mixture of normal distributions $N(-2, 1)$ with probability 0.7 and $N(2, 1)$ with probability 0.3, with $\text{E}(e_{ij}) = -0.8$ and $\text{Var}(e_{ij}) = 4.36$; $\sigma^2_b = 4.36$; $X_{ij} \sim \text{Uniform}(0, 1)$; and $L_{ij} \sim \text{Uniform}(0, 16)$.

5. $e_{ij}$ were drawn from a mixture of normal distributions $N(-2, 1)$ with probability 0.7 and $N(2, 1)$ with probability 0.3, with $\text{E}(e_{ij}) = -0.8$ and $\text{Var}(e_{ij}) = 4.36$; $\sigma^2_b = 4.36$; $X_{ij} \sim \text{Bernoulli}(0.5)$, where the covariate may be different for each individual in a cluster; and $L_{ij} \sim \text{Uniform}(0, 17)$.

6. $e_{ij}$ were drawn from a mixture of normal distributions $N(-2, 1)$ with probability 0.7 and $N(2, 1)$ with probability 0.3, with $\text{E}(e_{ij}) = -0.8$ and $\text{Var}(e_{ij}) = 4.36$; $\sigma^2_b = 4.36$; $X_{ij} \sim \text{Bernoulli}(0.5)$, where the covariate is same for each individual in a cluster; and $L_{ij} \sim \text{Uniform}(0, 17.5)$.

7. $\delta_{ij} \sim \text{Exp}(2.0)$ with $\text{E}(\log \delta_{ij}) = -1.270$ and $\text{Var}(\log \delta_{ij}) = 1.646$, $\sigma^2_b = 1.646$; $X_{ij} \sim \text{Uniform}(0, 1)$; and $L_{ij} \sim \text{Uniform}(0, 6)$.

8. $\delta_{ij} \sim \text{Exp}(2.0)$ with $\text{E}(\log \delta_{ij}) = -1.270$ and $\text{Var}(\log \delta_{ij}) = 1.646$, $\sigma^2_b = 1.646$; $X_{ij} \sim \text{Bernoulli}(0.5)$, where the covariate may be different for each individual in a cluster; and $L_{ij} \sim \text{Uniform}(0, 6.7)$. 

48
9. $\delta_{ij} \sim \text{Exp}(2.0)$ with $E(\log \delta_{ij}) = -1.270$ and $\text{Var}(\log \delta_{ij}) = 1.646$, $\sigma_b^2 = 1.646$; $X_{ij} \sim \text{Bernoulli}(0.5)$ where the covariate is same for each individual in a cluster; and $L_{ij} \sim \text{Uniform}(0, 6.7)$.

10. $\delta_{ij} \sim \text{Weibull}(2, 1)$, with $E(\log \delta_{ij}) = -0.289$ and $\text{Var}(\log \delta_{ij}) = 0.411$, $\sigma_b^2 = 0.411$; $X_{ij} \sim \text{Uniform}(0, 1)$; and $L_{ij} \sim \text{Uniform}(0, 10)$.

11. $\delta_{ij} \sim \text{Weibull}(2, 1)$, with $E(\log \delta_{ij}) = -0.289$ and $\text{Var}(\log \delta_{ij}) = 0.411$, $\sigma_b^2 = 0.411$; $X_{ij} \sim \text{Bernoulli}(0.5)$, where the covariate may be different for each individual in a cluster; and $L_{ij} \sim \text{Uniform}(0, 6)$.

12. $\delta_{ij} \sim \text{Weibull}(2, 1)$, with $E(\log \delta_{ij}) = -0.289$ and $\text{Var}(\log \delta_{ij}) = 0.411$, $\sigma_b^2 = 0.411$; $X_{ij} \sim \text{Bernoulli}(0.5)$ but the covariate is same for each individual in a cluster; and $L_{ij} \sim \text{Uniform}(0, 6)$.

13. $\delta_{ij} \sim \text{Gamma}(3, 1)$, with $E(\log \delta_{ij}) = 0.923$ and $\text{Var}(\log \delta_{ij}) = 0.395$, $\sigma_b^2 = 0.395$; $X_{ij} \sim \text{Uniform}(0, 1)$; and $L_{ij} \sim \text{Uniform}(0, 35)$.

14. This configuration is based on an interval censoring scenario with true $\beta = 4.0$. The $e_{ij}$ were drawn from a mixture of normal distributions $N(-2, 1)$ with probability 0.7 and $N(2, 1)$ with probability 0.3, with $E(e_{ij}) = -0.8$ and $\text{Var}(e_{ij}) = 4.36$; $\sigma_b^2 = 4.36$; $X_{ij} \sim \text{Bernoulli}(0.5)$, where the covariate may be different for each individual in a cluster. Five randomly chosen times from different lognormal distributions were chosen, $t_1 < t_2 < t_3 < t_4 < t_5$, in order to create a realistic interval censoring pattern without the constraint that the times are equally spaced. $t_1$ was generated from lognormal (-4.0, 0.16); $t_2$ was equal to $t_1$ plus another random number generated from lognormal distribution (-4.0, 0.16); $t_3$ was
equal to $t_2$ plus another random number generated from lognormal distribution (-4.0, 0.16); $t_4$ was equal to $t_3$ plus another random number generated from lognormal distribution (-4.0, 0.16); and $t_5$ was equal to $t_4$ plus another random number generated from lognormal distribution (-4.0, 0.16). Once the five times were set, six intervals were set, which are $[0, t_1], [t_1, t_2], [t_2, t_3], [t_3, t_4], [t_4, t_5]$ and $[t_5, \infty]$. We then generated $T_{ij}$ from our proposed cluster-specific AFT model, and could determine in which interval $T_{ij}$ lies. Using this scheme, we generated data with about 7% left censoring, 14% right censoring, and 79% interval censoring.

4.1.2 Starting Values

Optimization of loglikelihood requires good starting values for $\zeta$, as it can be very sensitive to the starting values. If the starting values are not chosen wisely, the MLEs may be not maximizing loglikelihood globally but only achieve a local maximum. Here we describe the approach we use to obtain starting values for the regression parameters.

Normal “Kernel”

For $K = 0$, we assume that the error comes from normal distribution and the data are not censored, and use SAS procedure proc mixed (SAS Institute, 2006) to obtain starting values for $\beta, \mu, \sigma_b$, and $\gamma$.

For $K = 1$, we use the same starting values for $\beta$ and $\sigma_b$ as with $K = 0$. We
choose 5 grid search points for $\psi$, from $-1.6$ to $1.6$ with step size $0.8$. For each $\psi$, we solve the following equations based on the expectation and variance of $\log T_{ij}$, where $T_{ij} = L_{ij}$ for right-censored data, and $T_{ij} = \frac{1}{2}(L_{ij} + R_{ij})$ for interval-censored data, and use the sample mean $\overline{\log(T_{ij})}$ and sample variance $s^2$ to approximate the expectation and variance to get the corresponding starting values of $\mu$ and $\gamma$:

$$\begin{align*}
\gamma &= \sqrt{\frac{s^2 - \sigma_b^2}{\text{Var}(Z_{ij})}} \\
\mu &= \log(T_{ij}) - \beta \bar{X} - \gamma \text{E}(Z_{ij}),
\end{align*}$$

(4.2)

where $\bar{X}$ is the sample mean of the covariate $X_{ij}$. The expectation and the variance of $Z_{ij}$ can be obtained from the SNP representation with the given "kernel" density and the tuning parameter $K$, evaluated at $\psi$.

For $K = 2$, we use a similar grid search method as $K = 1$ to obtain the starting value for each parameter. We choose 2 grid search points for $\psi_1$ and $\psi_2$, $-1.5$ and $1.5$. For each combination, $\psi = (\psi_1, \psi_2)^T$, we solve the above equations (4.2) to get the starting values for $\mu$ and $\gamma$.

For each grid search point for each $K$ and "kernel," we apply our proposed MCEM method to obtain MLEs and calculate the loglikelihood for each MLE. For each $K$, we choose as the MLE, the "solution" yielding the largest value of the loglikelihood.

**Exponential "Kernel"**

For $K = 0$, we assume that the error comes from a normal distribution and the data are not censored, as above, and use SAS procedure proc mixed to get the starting values for $\beta, \mu^*, \sigma_b$ and $\sigma^*_e$, where $\sigma^*_e$ is the variance of the error term and $\mu^*$ is the
intercept estimate. The starting value for $\gamma$ can be derived as
\[ \log\left(\sqrt{\sigma^*_e/1.646}\right), \]
and the starting value for $\mu$ can be derived as $\mu^* + 0.577e^\gamma$.

For $K = 1$, we use the same starting values for $\beta$, $\sigma_b$ and $\gamma$ as with $K = 0$. We perform a grid search over $\psi$, as in the normal “kernel” case, with the grid $-1.6$ to $1.6$ with step size $0.4$. For each $\psi$, we solve the following equation based on the expectation of $\log(T_{ij})$, where $T_{ij} = L_{ij}$ for right-censored data, and $T_{ij} = \frac{1}{2}(L_{ij} + R_{ij})$ for interval-censored data, and we use the sample mean $\overline{\log(T_{ij})}$ to approximate the expectation to get the corresponding starting value of $\mu$:
\[
\mu = \overline{\log(T_{ij})} - \beta \overline{X} - e^\gamma \mathbb{E}\{\log(Z_{ij})\},
\] (4.3)

where $\overline{X}$ is the sample mean of the covariate $X_{ij}$. The expectation of $\log Z_{ij}$ can be obtained from the SNP representation with the given “kernel” density and the tuning parameter $K$, evaluated at $\psi$.

For $K = 2$, we use a similar grid search method as $K = 1$ to obtain the starting value for each parameter. We choose grid search points for $\psi_1$ and $\psi_2$, from $-1.6$ and $1.6$ with step size $0.4$. For each combination, $\psi = (\psi_1, \psi_2)^T$, we solve the above equations (4.3) to get the starting value for $\mu$.

For each grid search point, we can calculate the loglikelihood and choose the one yielding the maximum likelihood and set the winner as the starting value. For this one set of starting values for each $K$, we apply our proposed MCEM to obtain the MLEs. It has been our experience that, for the exponential “kernel”, the optimization of the loglikelihood is not very sensitive to the starting values, so the way we choose starting values is a little bit different from that for the normal “kernel” as described.
above.

### 4.1.3 Simulation Results

For each scenario and each Monte Carlo dataset, we fitted model (3.2), (3.3) and (3.5) for \( K = 0, 1, 2 \). For convenience, we summarize the simulation results in Table 4.1. The table shows the number of data sets out of 100 for which each \( K \) and “kernel” combination was chosen by the HQ criterion. When error term was from a normal distribution, not surprisingly, the most frequently chosen density was the normal “kernel” with \( K = 0 \). When the error term was from a mixture of normal distributions, there are two modes for the true density, so \( K = 0 \) is not adequate to fit the model. From Table 4.1, \( K > 0 \) was chosen most often for mixture normal scenario. Therefore, when the true error distribution departs from normal distribution, that our proposed method can detect this departure is encouraging. When the true error comes from an exponential distribution or a Weibull distribution, in Section 3.8, we showed that the exponential “kernel” with \( K = 0 \) should fit perfectly. The results in Table 4.1 match this expectation, with the exponential “kernel” with \( K = 0 \) chosen most often for the exponential distribution and the Weibull distribution. The gamma distribution is somewhat like a normal and somewhat like a Weibull, so it is not surprising that, among 100 datasets, both exponential “kernel” and normal “kernel” were chosen. The approach we propose for detecting and representing departures from the exponential is encouraging.

We summarize the Monte Carlo average of the MLEs, the average of the estimated
standard error, the Monte Carlo empirical standard deviation, and coverage probability of 95% Wald confidence intervals (CI) in Tables 4.2 – 4.15. All the parameter estimates are almost unbiased. The estimated standard error is very close to the empirical standard deviation. All the coverage probabilities are around 95%. The average of estimated densities (of $e_{ij}$ or $\delta_{ij}$) chosen by HQ and true density (of $e_{ij}$ or $\delta_{ij}$) are shown in Figures 4.1 – 4.28. The 100 estimated densities chosen by HQ and true density are also shown in Figure 4.1 – 4.28. The figures show that the SNP density representation is capable of approximating the underlying true density well.
Table 4.1: Number of data sets choosing each $K$ and “kernel” combination.

<table>
<thead>
<tr>
<th>True Dist.</th>
<th>Cov.</th>
<th>Same</th>
<th>Censor</th>
<th>Normal Kernel</th>
<th>Exp Kernel</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>0 1 2</td>
<td>0 1 2</td>
<td></td>
</tr>
<tr>
<td>Normal*</td>
<td>Uni</td>
<td>N</td>
<td>R</td>
<td>98 2 0</td>
<td>0 0 0</td>
</tr>
<tr>
<td></td>
<td>Ber</td>
<td>N</td>
<td>R</td>
<td>94 2 3</td>
<td>0 0 1</td>
</tr>
<tr>
<td></td>
<td>Ber</td>
<td>Y</td>
<td>R</td>
<td>98 2 0</td>
<td>0 0 0</td>
</tr>
<tr>
<td>Mix Normal*</td>
<td>Uni</td>
<td>N</td>
<td>R</td>
<td>0 7 93</td>
<td>0 0 0</td>
</tr>
<tr>
<td></td>
<td>Ber</td>
<td>N</td>
<td>R</td>
<td>0 10 90</td>
<td>0 0 0</td>
</tr>
<tr>
<td></td>
<td>Ber</td>
<td>Y</td>
<td>R</td>
<td>0 11 89</td>
<td>0 0 0</td>
</tr>
<tr>
<td>Exp**</td>
<td>Uni</td>
<td>N</td>
<td>R</td>
<td>0 0 0</td>
<td>96 4 0</td>
</tr>
<tr>
<td></td>
<td>Ber</td>
<td>N</td>
<td>R</td>
<td>0 0 0</td>
<td>96 2 2</td>
</tr>
<tr>
<td></td>
<td>Ber</td>
<td>Y</td>
<td>R</td>
<td>0 0 0</td>
<td>95 4 1</td>
</tr>
<tr>
<td>Weibull**</td>
<td>Uni</td>
<td>N</td>
<td>R</td>
<td>0 0 0</td>
<td>96 2 2</td>
</tr>
<tr>
<td></td>
<td>Ber</td>
<td>N</td>
<td>R</td>
<td>0 0 1</td>
<td>98 1 0</td>
</tr>
<tr>
<td></td>
<td>Ber</td>
<td>Y</td>
<td>R</td>
<td>0 0 0</td>
<td>94 4 2</td>
</tr>
<tr>
<td>Gamma**</td>
<td>Uni</td>
<td>N</td>
<td>R</td>
<td>13 6 0</td>
<td>52 1 28</td>
</tr>
<tr>
<td>Mix Normal*</td>
<td>Ber</td>
<td>N</td>
<td>I</td>
<td>2 95 3</td>
<td>0 0 0</td>
</tr>
</tbody>
</table>

Where * is the true distribution of $e_{ij}$ and ** is the true distribution of $\delta_{ij}$; Cov. is the true distribution of the covariate, which can either be Uniform (Uni) or Bernoulli (Ber) distribution; Same is whether the covariate is same for each individual in a cluster, where Y means same and N means different; Censor specifies the censoring pattern either right-censored (R) or interval-censored (I); $K$ is the tuning parameter, which can be 0, 1, 2.
Table 4.2: Right-censored data with normal error term and Uniform(0,1) covariate.

<table>
<thead>
<tr>
<th>Parameter (True)</th>
<th>Ave.</th>
<th>Est. SE</th>
<th>Emp. SD</th>
<th>Cov. Prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta$ (2.000)</td>
<td>2.034</td>
<td>0.246</td>
<td>0.232</td>
<td>0.94</td>
</tr>
<tr>
<td>$\sigma$ (1.414)</td>
<td>1.419</td>
<td>0.133</td>
<td>0.129</td>
<td>0.97</td>
</tr>
<tr>
<td>$E(e_{ij})$ (-1.500)</td>
<td>-1.529</td>
<td>0.198</td>
<td>0.206</td>
<td>0.93</td>
</tr>
<tr>
<td>$sd(e_{ij})$ (1.414)</td>
<td>1.416</td>
<td>0.057</td>
<td>0.056</td>
<td>0.95</td>
</tr>
</tbody>
</table>

Parameter (True) is the true value for each parameter; $E(e_{ij})$ is the expectation of $e_{ij}$; $sd(e_{ij})$ is the standard deviation of $e_{ij}$; Ave. is the average of the estimates over 100 data sets; Estimated SE is the average of the estimated standard errors; Emp. SD is the empirical standard deviation; Cov. Prob. is the coverage probability of 95% Wald confidence intervals (CI).

Figure 4.1: (a) Average of 100 estimated densities chosen by HQ (dash line) and true density of $e_{ij}$ (solid line). (b) 100 estimated densities chosen by HQ and true density (solid line): normal scenario with Uniform(0,1) covariate.
Figure 4.2: (a) Average of 100 estimated densities chosen by HQ (dash line) and true density of $\delta_{ij}$ (solid line). (b) 100 estimated densities chosen by HQ and true density (solid line): normal scenario with Uniform(0,1) covariate.
Table 4.3: Right-censored data with normal error term and different Bernoulli(0.5) covariate in each cluster. Table entries are same as in Table 4.2.

<table>
<thead>
<tr>
<th>Parameter (True)</th>
<th>Ave.</th>
<th>Est. SE</th>
<th>Emp. SD</th>
<th>Cov. Prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta$ (2.000)</td>
<td>2.024</td>
<td>0.142</td>
<td>0.140</td>
<td>0.94</td>
</tr>
<tr>
<td>$\sigma_b$ (1.414)</td>
<td>1.419</td>
<td>0.134</td>
<td>0.119</td>
<td>0.98</td>
</tr>
<tr>
<td>$\text{E}(e_{ij})$ ($-1.500$)</td>
<td>$-1.529$</td>
<td>0.172</td>
<td>0.169</td>
<td>0.94</td>
</tr>
<tr>
<td>$\text{sd}(e_{ij})$ (1.414)</td>
<td>1.413</td>
<td>0.056</td>
<td>0.053</td>
<td>0.95</td>
</tr>
</tbody>
</table>

Figure 4.3: (a) Average of 100 estimated densities chosen by HQ (dashed line) and true density of $e_{ij}$ (solid line). (b) 100 estimated densities chosen by HQ and true density (solid line): normal scenario with different Bernoulli(0.5) covariate in each cluster.
Figure 4.4: (a) Average of 100 estimated densities chosen by HQ (dash line) and true density of $\delta_{ij}$ (solid line). (b) 100 estimated densities chosen by HQ and true density (solid line): normal scenario with different Bernoulli(0.5) covariate in each cluster.
Table 4.4: Right-censored data with normal error term and same Bernoulli(0.5) covariate in each cluster. Table entries are same as in Table 4.2.

<table>
<thead>
<tr>
<th>Parameter (True)</th>
<th>Ave.</th>
<th>Est. SE</th>
<th>Emp. SD</th>
<th>Cov. Prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta$ (2.000)</td>
<td>1.996</td>
<td>0.315</td>
<td>0.353</td>
<td>0.95</td>
</tr>
<tr>
<td>$\sigma_b$ (1.414)</td>
<td>1.381</td>
<td>0.131</td>
<td>0.131</td>
<td>0.90</td>
</tr>
<tr>
<td>$E(e_{ij})$ ($-1.500$)</td>
<td>$-1.520$</td>
<td>0.222</td>
<td>0.241</td>
<td>0.95</td>
</tr>
<tr>
<td>$sd(e_{ij})$ (1.414)</td>
<td>1.408</td>
<td>0.059</td>
<td>0.067</td>
<td>0.91</td>
</tr>
</tbody>
</table>

Figure 4.5: (a) Average of 100 estimated densities chosen by HQ (dash line) and true density of $e_{ij}$ (solid line). (b) 100 estimated densities chosen by HQ and true density (solid line): normal scenario with same Bernoulli(0.5) covariate in each cluster.
Figure 4.6: (a) Average of 100 estimated densities chosen by HQ (dash line) and true density of $\delta_{ij}$ (solid line). (b) 100 estimated densities chosen by HQ and true density (solid line): normal scenario with same Bernoulli(0.5) covariate in each cluster.
Table 4.5: Right-censored data with mixture of normal error term and Uniform(0,1) covariate. Table entries are same as in Table 4.2.

<table>
<thead>
<tr>
<th>Parameter (True)</th>
<th>Ave.</th>
<th>Est. SE</th>
<th>Emp. SD</th>
<th>Cov. Prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta$ (2.000)</td>
<td>2.012</td>
<td>0.242</td>
<td>0.250</td>
<td>0.95</td>
</tr>
<tr>
<td>$\sigma_0$ (2.089)</td>
<td>2.066</td>
<td>0.171</td>
<td>0.182</td>
<td>0.92</td>
</tr>
<tr>
<td>$E(e_{ij})$ ($-0.800$)</td>
<td>$-0.830$</td>
<td>0.260</td>
<td>0.303</td>
<td>0.92</td>
</tr>
<tr>
<td>$sd(e_{ij})$ (2.089)</td>
<td>2.093</td>
<td>0.077</td>
<td>0.071</td>
<td>0.97</td>
</tr>
</tbody>
</table>

Figure 4.7: (a) Average of 100 estimated densities chosen by HQ (dash line) and true density of $e_{ij}$ (solid line). (b) 100 estimated densities chosen by HQ and true density (solid line): mixture of normal scenario with Uniform(0,1) covariate.
Figure 4.8: (a) Average of 100 estimated densities chosen by HQ (dash line) and true density of $\delta_{ij}$ (solid line). (b) 100 estimated densities chosen by HQ and true density (solid line): mixture of normal scenario with Uniform(0,1) covariate.
Table 4.6: Right-censored data with mixture of normal error term and different Bernoulli(0.5) covariate in each cluster. Table entries are same as in Table 4.2.

<table>
<thead>
<tr>
<th>Parameter (True)</th>
<th>Ave.</th>
<th>Est. SE</th>
<th>Emp. SD</th>
<th>Cov. Prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta$ (2.000)</td>
<td>2.029</td>
<td>0.137</td>
<td>0.135</td>
<td>0.97</td>
</tr>
<tr>
<td>$\sigma_b$ (2.089)</td>
<td>2.079</td>
<td>0.173</td>
<td>0.174</td>
<td>0.94</td>
</tr>
<tr>
<td>$\mathbb{E}(e_{ij})$ ($-0.800$)</td>
<td>$-0.845$</td>
<td>0.224</td>
<td>0.259</td>
<td>0.94</td>
</tr>
<tr>
<td>$\text{sd}(e_{ij})$ (2.089)</td>
<td>2.091</td>
<td>0.077</td>
<td>0.071</td>
<td>0.95</td>
</tr>
</tbody>
</table>

Figure 4.9: (a) Average of 100 estimated densities chosen by HQ (dash line) and true density of $e_{ij}$ (solid line). (b) 100 estimated densities chosen by HQ and true density (solid line): mixture of normal scenario with different Bernoulli(0.5) covariate in each cluster.
Figure 4.10: (a) Average of 100 estimated densities chosen by HQ (dash line) and true density of $\delta_{ij}$ (solid line). (b) 100 estimated densities chosen by HQ and true density (solid line): mixture of normal scenario with different Bernoulli(0.5) covariate in each cluster.
Table 4.7: Right-censored data with mixture of normal error term and same Bernoulli(0.5) covariate in each cluster. Table entries are same as in Table 4.2.

<table>
<thead>
<tr>
<th>Parameter (True)</th>
<th>Ave.</th>
<th>Est. SE</th>
<th>Emp. SD</th>
<th>Cov. Prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta$ (2.000)</td>
<td>1.946</td>
<td>0.444</td>
<td>0.453</td>
<td>0.94</td>
</tr>
<tr>
<td>$\sigma_b$ (2.089)</td>
<td>2.056</td>
<td>0.176</td>
<td>0.150</td>
<td>0.98</td>
</tr>
<tr>
<td>$E(e_{ij})$ (-0.800)</td>
<td>-0.812</td>
<td>0.322</td>
<td>0.341</td>
<td>0.94</td>
</tr>
<tr>
<td>$sd(e_{ij})$ (2.089)</td>
<td>2.095</td>
<td>0.076</td>
<td>0.078</td>
<td>0.94</td>
</tr>
</tbody>
</table>

Figure 4.11: (a) Average of 100 estimated densities chosen by HQ (dash line) and true density of $e_{ij}$ (solid line). (b) 100 estimated densities chosen by HQ and true density (solid line): mixture normal scenario with same Bernoulli(0.5) covariate in each cluster.
Figure 4.12: (a) Average of 100 estimated densities chosen by HQ (dash line) and true density of $\delta_{ij}$ (solid line). (b) 100 estimated densities chosen by HQ and true density (solid line): mixture normal scenario with same Bernoulli(0.5) covariate in each cluster.
Table 4.8: Right-censored data with exponential error term and Uniform(0,1) covariate. Table entries are same as in Table 4.2.

<table>
<thead>
<tr>
<th>Parameter (True)</th>
<th>Ave.</th>
<th>Est. SE</th>
<th>Emp. SD</th>
<th>Cov. Prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta$ (2.000)</td>
<td>2.049</td>
<td>0.205</td>
<td>0.199</td>
<td>0.93</td>
</tr>
<tr>
<td>$\sigma_0$ (1.283)</td>
<td>1.269</td>
<td>0.116</td>
<td>0.113</td>
<td>0.91</td>
</tr>
<tr>
<td>$E(e_{ij})$ (-1.270)</td>
<td>-1.301</td>
<td>0.170</td>
<td>0.177</td>
<td>0.94</td>
</tr>
<tr>
<td>$sd(e_{ij})$ (1.283)</td>
<td>1.291</td>
<td>0.057</td>
<td>0.057</td>
<td>0.95</td>
</tr>
</tbody>
</table>

Figure 4.13: (a) Average of 100 estimated densities chosen by HQ (dash line) and true density of $\delta_{ij}$ (solid line). (b) 100 estimated densities chosen by HQ and true density (solid line): exponential scenario with Uniform(0,1) covariate.
Figure 4.14: (a) Average of 100 estimated densities chosen by HQ (dash line) and true density of $e_{ij}$ (solid line). (b) 100 estimated densities chosen by HQ and true density (solid line): exponential scenario with Uniform(0,1) covariate.
Table 4.9: Right-censored data with exponential error term and different Bernoulli(0.5) covariate in each cluster. Table entries are same as in Table 4.2.

<table>
<thead>
<tr>
<th>Parameter (True)</th>
<th>Ave.</th>
<th>Est. SE</th>
<th>Emp. SD</th>
<th>Cov. Prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \beta ) (2.000)</td>
<td>2.022</td>
<td>0.123</td>
<td>0.119</td>
<td>0.96</td>
</tr>
<tr>
<td>( \sigma_0 ) (1.283)</td>
<td>1.269</td>
<td>0.116</td>
<td>0.110</td>
<td>0.94</td>
</tr>
<tr>
<td>( E(e_{ij}) ) (-1.270)</td>
<td>-1.295</td>
<td>0.152</td>
<td>0.161</td>
<td>0.95</td>
</tr>
<tr>
<td>( sd(e_{ij}) ) (1.283)</td>
<td>1.289</td>
<td>0.058</td>
<td>0.053</td>
<td>0.96</td>
</tr>
</tbody>
</table>

Figure 4.15: (a) Average of 100 estimated densities chosen by HQ (dash line) and true density of \( \delta_{ij} \) (solid line). (b) 100 estimated densities chosen by HQ and true density (solid line): exponential scenario with different Bernoulli(0.5) covariate in each cluster.
Figure 4.16: (a) Average of 100 estimated densities chosen by HQ (dash line) and true density of \( e_{ij} \) (solid line). (b) 100 estimated densities chosen by HQ and true density (solid line): exponential scenario with different Bernoulli(0.5) covariate in each cluster.
Table 4.10: Right-censored data with exponential error term and same Bernoulli(0.5) covariate in each cluster. Table entries are same as in Table 4.2.

<table>
<thead>
<tr>
<th>Parameter (True)</th>
<th>Ave.</th>
<th>Est. SE</th>
<th>Emp. SD</th>
<th>Cov. Prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta$ (2.000)</td>
<td>2.078</td>
<td>0.292</td>
<td>0.300</td>
<td>0.95</td>
</tr>
<tr>
<td>$\sigma_b$ (1.283)</td>
<td>1.289</td>
<td>0.121</td>
<td>0.122</td>
<td>0.93</td>
</tr>
<tr>
<td>$E(e_{ij})$ (-1.270)</td>
<td>-1.278</td>
<td>0.202</td>
<td>0.211</td>
<td>0.95</td>
</tr>
<tr>
<td>sd($e_{ij}$) (1.283)</td>
<td>1.293</td>
<td>0.057</td>
<td>0.058</td>
<td>0.94</td>
</tr>
</tbody>
</table>

Figure 4.17: (a) Average of 100 estimated densities chosen by HQ (dash line) and true density of $\delta_{ij}$ (solid line). (b) 100 estimated densities chosen by HQ and true density (solid line): Exponential scenario with same Bernoulli(0.5) covariate in each cluster.
Figure 4.18: (a) Average of 100 estimated densities chosen by HQ (dash line) and true density of \( e_{ij} \) (solid line). (b) 100 estimated densities chosen by HQ and true density (solid line): Exponential scenario with same Bernoulli(0.5) covariate in each cluster.
Table 4.11: Right-censored data with Weibull error term and Uniform(0,1) covariate. Table entries are same as in Table 4.2.

<table>
<thead>
<tr>
<th>Parameter (True)</th>
<th>Ave.</th>
<th>Est. SE</th>
<th>Emp. SD</th>
<th>Cov. Prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta$ (2.000)</td>
<td>1.998</td>
<td>0.106</td>
<td>0.096</td>
<td>0.99</td>
</tr>
<tr>
<td>$\sigma_b$ (0.641)</td>
<td>0.632</td>
<td>0.056</td>
<td>0.050</td>
<td>0.94</td>
</tr>
<tr>
<td>$E(e_{ij})$ (-0.289)</td>
<td>-0.294</td>
<td>0.110</td>
<td>0.097</td>
<td>0.90</td>
</tr>
<tr>
<td>$sd(e_{ij})$ (0.641)</td>
<td>0.637</td>
<td>0.040</td>
<td>0.030</td>
<td>0.94</td>
</tr>
</tbody>
</table>

Figure 4.19: (a) Average of 100 estimated densities chosen by HQ (dash line) and true density of $\delta_{ij}$ (solid line); (b) 100 estimated densities chosen by HQ and true density (solid line): Weibull scenario with Uniform(0,1) covariate.
Figure 4.20: (a) Average of 100 estimated densities chosen by HQ (dash line) and true density of $e_{ij}$ (solid line); (b) 100 estimated densities chosen by HQ and true density (solid line): Weibull scenario with Uniform(0,1) covariate.
Table 4.12: Right-censored data with Weibull error term and different Bernoulli(0.5) covariate in each cluster. Table entries are same as in Table 4.2.

<table>
<thead>
<tr>
<th>Parameter (True)</th>
<th>Ave.</th>
<th>Est. SE</th>
<th>Emp. SD</th>
<th>Cov. Prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta$ (2.000)</td>
<td>1.996</td>
<td>0.063</td>
<td>0.057</td>
<td>0.94</td>
</tr>
<tr>
<td>$\sigma_0$ (0.641)</td>
<td>0.630</td>
<td>0.057</td>
<td>0.052</td>
<td>0.96</td>
</tr>
<tr>
<td>$E(e_{ij})$ (-0.289)</td>
<td>-0.299</td>
<td>0.075</td>
<td>0.079</td>
<td>0.93</td>
</tr>
<tr>
<td>$sd(e_{ij})$ (0.641)</td>
<td>0.639</td>
<td>0.028</td>
<td>0.027</td>
<td>0.94</td>
</tr>
</tbody>
</table>

Figure 4.21: (a) Average of 100 estimated densities chosen by HQ (dash line) and true density of $\delta_{ij}$ (solid line); (b) 100 estimated densities chosen by HQ and true density (solid line): Weibull scenario with different Bernoulli(0.5) covariate in each cluster.
Figure 4.22: (a) Average of 100 estimated densities chosen by HQ (dash line) and true density of $e_{ij}$ (solid line); (b) 100 estimated densities chosen by HQ and true density (solid line): Weibull scenario with different Bernoulli(0.5) covariate in each cluster.
Table 4.13: Right-censored data with Weibull error and same Bernoulli(0.5) covariate in each cluster. Table entries are same as in Table 4.2.

<table>
<thead>
<tr>
<th>Parameter (True)</th>
<th>Ave.</th>
<th>Est. SE</th>
<th>Emp. SD</th>
<th>Cov. Prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>β (2.000)</td>
<td>2.003</td>
<td>0.145</td>
<td>0.145</td>
<td>0.96</td>
</tr>
<tr>
<td>σb (0.641)</td>
<td>0.638</td>
<td>0.059</td>
<td>0.056</td>
<td>0.97</td>
</tr>
<tr>
<td>E(e_{ij}) (-0.289)</td>
<td>-0.285</td>
<td>0.100</td>
<td>0.112</td>
<td>0.89</td>
</tr>
<tr>
<td>sd(e_{ij}) (0.641)</td>
<td>0.639</td>
<td>0.028</td>
<td>0.026</td>
<td>0.95</td>
</tr>
</tbody>
</table>

Figure 4.23: (a) Average of 100 estimated densities chosen by HQ (dash line) and true density of δ_{ij} (solid line); (b) 100 estimated densities chosen by HQ and true density (solid line): Weibull scenario with same Bernoulli(0.5) covariate in each cluster.
Figure 4.24: (a) Average of 100 estimated densities chosen by HQ (dash line) and true density of \( e_{ij} \) (solid line); (b) 100 estimated densities chosen by HQ and true density (solid line): Weibull scenario with same Bernoulli(0.5) covariate in each cluster.
Table 4.14: Right-censored data with Gamma error term and Uniform(0,1) covariate. Table entries are same as in Table 4.2

<table>
<thead>
<tr>
<th>Parameter (True)</th>
<th>Ave.</th>
<th>Est. SE</th>
<th>Emp. SD</th>
<th>Cov. Prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta$ (2.000)</td>
<td>2.013</td>
<td>0.110</td>
<td>0.103</td>
<td>0.98</td>
</tr>
<tr>
<td>$\sigma_b$ (0.628)</td>
<td>0.625</td>
<td>0.057</td>
<td>0.053</td>
<td>0.93</td>
</tr>
<tr>
<td>$E(e_{ij})$ (0.923)</td>
<td>0.910</td>
<td>0.088</td>
<td>0.089</td>
<td>0.93</td>
</tr>
<tr>
<td>$sd(e_{ij})$ (0.628)</td>
<td>0.640</td>
<td>0.032</td>
<td>0.029</td>
<td>0.93</td>
</tr>
</tbody>
</table>

Figure 4.25: (a) Average of 100 estimated densities chosen by HQ (dash line) and true density of $\delta_{ij}$ (solid line); (b) 100 estimated densities chosen by HQ and true density (solid line): Gamma scenario with Uniform(0,1) covariate.
Figure 4.26: (a) Average of 100 estimated densities chosen by HQ (dash line) and true density of \( e_{ij} \) (solid line); (b) 100 estimated densities chosen by HQ and true density (solid line): Gamma scenario with Uniform(0,1) covariate.
Table 4.15: Interval censored data with mixture of normal error term and different Bernoulli(0.5) covariate in each cluster. Table entries are same in Table 4.2.

<table>
<thead>
<tr>
<th>Parameter (True)</th>
<th>Ave.</th>
<th>Est. SE</th>
<th>Emp. SD</th>
<th>Cov. Prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta$ (4.000)</td>
<td>4.009</td>
<td>0.242</td>
<td>0.235</td>
<td>0.97</td>
</tr>
<tr>
<td>$\sigma_0$ (2.089)</td>
<td>2.047</td>
<td>0.195</td>
<td>0.178</td>
<td>0.93</td>
</tr>
<tr>
<td>$E(e_{ij}) (-0.800)$</td>
<td>-0.805</td>
<td>0.265</td>
<td>0.243</td>
<td>0.98</td>
</tr>
<tr>
<td>$sd(e_{ij}) (2.089)$</td>
<td>2.089</td>
<td>0.104</td>
<td>0.101</td>
<td>0.97</td>
</tr>
</tbody>
</table>

Figure 4.27: (a) Average of 100 estimated densities chosen by HQ (dash line) and true density of $e_{ij}$ (solid line); (b) 100 estimated densities chosen by HQ and true density (solid line): Interval censored mixture of normal scenario with different Bernoulli(0.5) covariate.
Figure 4.28: (a) Average of 100 estimated densities chosen by HQ (dash line) and true density of $\delta_{ij}$ (solid line) and ; (b) 100 estimated densities chosen by HQ and true density (solid line): Interval censored mixture of normal scenario with different Bernoulli(0.5) covariate.
4.2 Hypothesis Testing for Variance Components

We discussed hypothesis testing for variance components in Section 3.7.2. We conducted simulation studies to investigate if the nominal size of the proposed likelihood ratio test is close to 0.05 by using the level-\( \alpha = 0.05 \) likelihood ratio test to test the hypothesis

\[ H_0 : \sigma_b^2 = 0 \text{ vs. } H_0 : \sigma_b^2 > 0. \]

We simulated 100 Monte Carlo data sets. In each of 100 data sets, we generated \( T_{ij}, i = 1, \ldots, 100, j = 1, \ldots, 6 \), from the cluster-specific AFT (3.2) model under \( H_0 : \sigma_b^2 = 0 \), i.e., under

\[ \log(T_{ij}) = \beta X_{ij} + e_{ij}, \tag{4.4} \]

where the true \( \beta = 2.0 \); and covariate \( X_{ij} \) was always from Uniform(0,1), which may be different for the individuals in a cluster. We present simulation results under the following scenarios:

1. \( e_{ij} \sim N(\mu, \sigma_{e}^2) \) with \( \mu = -1.5, \sigma_{e}^2 = 2.0 \); \( X_{ij} \sim \text{Uniform}(0,1) \); and \( L_{ij} \sim \text{Uniform}(0,4) \).

2. \( e_{ij} \) were drawn from a mixture of normal distributions \( N(-2, 1) \) with probability 0.7 and \( N(2, 1) \) with probability 0.3, with \( \text{E}(e_{ij}) = -0.8 \) and \( \text{Var}(e_{ij}) = 4.36 \); \( X_{ij} \sim \text{Uniform}(0,1) \); and \( L_{ij} \sim \text{Uniform}(0,11) \).

3. \( \delta_{ij} \sim \text{Exp}(2.0) \) with \( \text{E}(\log \delta_{ij}) = -1.270 \) and \( \text{Var}(\log \delta_{ij}) = 1.646 \); \( X_{ij} \sim \text{Uniform}(0,1) \); and \( L_{ij} \sim \text{Uniform}(0,4.7) \).
4. $\delta_{ij}$ \sim \text{Weibull}(2, 1)$, with $E(\log \delta_{ij}) = -0.289$ and $\text{Var}(\log \delta_{ij}) = 0.411$; $X_{ij} \sim \text{Uniform}(0, 1)$; and $L_{ij} \sim \text{Uniform}(0, 9.2)$.

As in Section 3.7.2, we first fit the “full” model with $\sigma_b^2$ estimated, determine $K$ and “kernel.” Then we fit the “reduced” model with $\sigma_b^2 = 0$ using same $K$ and “kernel.” In the “reduced” model, the density of the error term is thus still approximated by the SNP representation. Based on these fits, we calculated the likelihood ratio test statistic and compared it to the critical point of a mixture of $\chi^2(0)$ and $\chi^2(1)$ with equal probability 0.5 for a 0.05 level test to determine if we should reject the null hypothesis or not. Table 4.16 shows the Monte Carlo rejection rates for each simulation scenario and shows that the test appears to respect the nominal level.

<table>
<thead>
<tr>
<th>True Distribution</th>
<th>Nominal Level</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lognormal</td>
<td>0.04</td>
</tr>
<tr>
<td>Log Mixture of Normal</td>
<td>0.05</td>
</tr>
<tr>
<td>Exponential</td>
<td>0.04</td>
</tr>
<tr>
<td>Weibull</td>
<td>0.04</td>
</tr>
</tbody>
</table>

Table 4.16: Nominal level for variance components hypothesis testing

True Distribution is the distribution of $\delta_{ij}$; Nominal Level is $P(LRT \geq C)$ under $H_0$, where $C$ is the critical point of mixture of $\chi^2(0)$ and $\chi^2(1)$ with equal probability 0.5.

### 4.3 Violation of Normality of Random Effects

In our proposed cluster-specific AFT model, we assume that the random effect $b_i$ has a normal distribution with mean zero and variance $\sigma_b^2$. However, if the random effect departs from a normal distribution, it is important to evaluate how the MLEs and variance components estimator will be affected. We show one simulation result
with right-censored data addressing this issue here. We simulated 100 Monte Carlo data sets. In each dataset, we generated $T_{ij}$, $i = 1, \ldots, 100$, $j = 1, \ldots, 6$, from the cluster-specific AFT model

$$\log(T_{ij}) = \beta X_{ij} + b_i + e_{ij},$$

where true $\beta = 2.0$; covariate $X_{ij}$ was from Uniform(0, 1), which may be different for each individual in a cluster; the censoring time $L_{ij}$ was drawn from Uniform(0, 28); $e_{ij}$ was drawn from a mixture of normal distributions $N(-0.9, 1)$ with probability 0.7 and $N(2.1, 1)$ with probability 0.3; and the random effect $b_i$ comes from same mixture of normal distributions as the error term.

From Table 4.17, the average of estimates is very close to the truth for each parameter. The average of estimated standard errors is very close to the empirical standard deviation and coverage probability is around 95%. Even though the random effect is not from a normal distribution, the MLEs still show good operating characteristics. The average of the estimated error densities in Figure 4.29 is very close to the true error density. This simulation shows that, even though the distribution of the random effects departs from a normal distribution, we obtain unbiased estimator for each parameter and a good density estimate for the error term. A deeper study of this issue is a subject for future research.
Table 4.17: Right-censored data with mixture of normal error term, Uniform(0,1) covariate, and mixture of normal random effect. Table entries are same in Table 4.2.

<table>
<thead>
<tr>
<th>Parameter (True)</th>
<th>Ave.</th>
<th>Est. SE</th>
<th>Emp. SD</th>
<th>Cov. Prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta$ (2.000)</td>
<td>2.050</td>
<td>0.249</td>
<td>0.235</td>
<td>0.95</td>
</tr>
<tr>
<td>$\sigma_0$ (1.700)</td>
<td>1.663</td>
<td>0.167</td>
<td>0.131</td>
<td>0.96</td>
</tr>
<tr>
<td>$E(e_{ij})$ (0.000)</td>
<td>$-0.044$</td>
<td>0.231</td>
<td>0.231</td>
<td>0.96</td>
</tr>
<tr>
<td>$sd(e_{ij})$ (1.700)</td>
<td>1.708</td>
<td>0.069</td>
<td>0.064</td>
<td>0.98</td>
</tr>
</tbody>
</table>

Figure 4.29: (a) Average of 100 estimated densities chosen by HQ (dash line) and true density of $e_{ij}$ (solid line); (b) 100 estimated densities chosen by HQ and true density of $e_{ij}$ (solid line): Relaxing the distribution of random effect.
4.4 Applications

We consider three applications to illustrate our proposed methods.

4.4.1 Litter-matched Tumorigenesis Study

One application is from the litter-matched tumorigenesis study (Mantel et al., 1977) introduced in Section 1.2. There were three rats in each 50 female litters. In each litter, one rat was treated while the others were not. The failure time is time to tumor appearance and the censoring time is time to death. For the $i$th litter, the treatment indicator $X_{ij} = 0$ if the $j$th rat is drug-treated, and $X_{ij} = 1$ otherwise, where $i = 1, \ldots, 50$, $j = 1, 2, 3$. About 74% of time to tumor appearance was censored by death. We are interested in assessing whether or not the time to tumor appearance for the control group tends to be longer than that for drug-treated group. Because the failure times may be correlated in each cluster, we fitted the cluster-specific AFT model. Table 4.18 shows the MLE, estimated standard error, and 95% Wald confidence interval (CI) for each parameter based on the preferred (by HQ) SNP representation with $K = 0$ and exponential “kernel.”

Our estimate for Table 4.18: Inference results for litter-matched tumorigenesis study data set

<table>
<thead>
<tr>
<th>Parameter</th>
<th>MLE</th>
<th>Est. SE</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta$</td>
<td>0.206</td>
<td>0.093</td>
<td>(0.024, 0.388)</td>
</tr>
<tr>
<td>$\sigma_b$</td>
<td>0.116</td>
<td>0.134</td>
<td>(-0.147, 0.379)</td>
</tr>
<tr>
<td>$E(e_{ij})$</td>
<td>4.629</td>
<td>0.068</td>
<td>(4.496, 4.762)</td>
</tr>
<tr>
<td>$sd(e_{ij})$</td>
<td>0.321</td>
<td>0.044</td>
<td>(0.235, 0.407)</td>
</tr>
</tbody>
</table>

Parameter is parameter we are most interest of; $E(e_{ij})$ is the expectation of $e_{ij}$; $sd(e_{ij})$ is the standard deviation of $e_{ij}$; MLE is maximum likelihood estimate; Est. SE is the estimated standard errors; 95% CI is the 95% Wald confidence intervals (CI).
\( \beta \) is close to the values of 0.165 and 0.168 produced from rank with logrank score method, the Buckley-James method, and independence with logrank score methods in Lee et al. (1993). But it is not close to the value of 0.10 produced by Pan et al. (2000), who assumed normal distributions for both the random effect and error term. In contrast, Lee et al. (1993) did not make any assumption about the distribution of the error term. If we assume the error term comes from normal distribution, the estimate of density is shown as the dashed line in Figure 4.30, while the solid line is the estimated density for our final chosen model, and shows that normal assumption for the error term in Pan et al. (2000) is not appropriate.

![Figure 4.30](image)

Figure 4.30: Dashed line is the density estimate based on normal assumption; solid line is the density estimate of our final chosen model.

The 95% confidence interval for \( \beta \) does not cover zero, indicating that we have enough evidence to reject the null hypothesis \( H_0 : \beta = 0 \). We can conclude that
the time to tumor appearance for the control group is about \( e^{0.206} - 1 = 0.229 \) times longer than the drug-treated group on average.

We are also interested in whether the failure times within clusters are correlated or not. We carried out hypothesis test \( H_0 : \sigma_b^2 = 0 \) vs. \( H_a : \sigma_b^2 > 0 \). The likelihood ratio test statistic is equal to 0.97, which is less than 95\% critical point of mixture chi-square of 0 and 1 with equal probability 0.5. We thus conclude that there is no strong evidence to support that the failure times in each litter are correlated. If we assume that there is no correlation in each cluster, removing the random effect from our model and still using the SNP representation for the error term, we obtain an estimate of \( \beta \) 0.207, which is very close to 0.206.

After the final model is chosen, using the method in Section 2.3, we can generate a sample of the failure times based on the distribution of random effect and error term for a given setting of the covariate. We generate 150000 failure times for treated rats. Figure 4.31 shows the histogram of the failure time, the estimate of smooth density of the failure time, survival function, and hazard function.

![Figure 4.31](a) Histogram of failure time and estimate of density; (b) Survival function; (c) Hazard function
4.4.2 Diabetic Retinopathy Study

Our second example is Diabetic Retinopathy Study (DRS), also discussed in Section 1.3, a randomized trial conducted by the National Eye Institute to evaluate the photocoagulation treatment for proliferative diabetic retinopathy. There were 197 patients, where photocoagulation was assigned to one eye for each patient and the other eye as control. One main objective was to assess whether the time to occurrence of severe visual loss for the control eye is longer than the treated eye. The failure times to occurrence of severe visual loss clearly may be correlated in each person. For the \( i \)th patient, the treatment indicator \( X_{ij} = 1 \) if the \( j \)th eye is photocoagulation-treated, \( X_{ij} = 0 \) otherwise, where \( i = 1, \ldots, 197, j = 1, 2 \). We fitted the cluster-specific AFT model with treatment indicator as the only covariate. The final model chosen by the HQ criterion was the normal “kernel” with \( K = 0 \), suggesting that the error term comes from a normal distribution. Table 4.19 shows the MLE, estimated standard error and 95% Wald confidence interval for each parameter. The confidence interval for \( \beta \) does not cover zero, so we have strong evidence to reject \( H_0 : \beta = 0 \). We conclude that time to occurrence of severe visual loss is about \( e^{1.016} - 1 = 1.762 \) times longer for photocoagulation treated eye than the control eye on average.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>MLE</th>
<th>Est. SE</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \beta )</td>
<td>1.016</td>
<td>0.199</td>
<td>(0.626, 1.406)</td>
</tr>
<tr>
<td>( \sigma_b )</td>
<td>1.103</td>
<td>0.195</td>
<td>(0.721, 1.485)</td>
</tr>
<tr>
<td>( E(e_{ij}) )</td>
<td>3.824</td>
<td>0.160</td>
<td>(3.510, 4.138)</td>
</tr>
<tr>
<td>( se(e_{ij}) )</td>
<td>1.530</td>
<td>0.116</td>
<td>(1.303, 1.757)</td>
</tr>
</tbody>
</table>
We are also interested in whether the failure times in each cluster are correlated or not. We carried out the hypothesis test $H_0 : \sigma_b^2 = 0$ vs. $H_a : \sigma_b^2 > 0$. The LRT test statistic is equal to 13.59, which is greater than 95% critical point of a mixture of $\chi^2(0)$ and $\chi^2(1)$ with equal probability 0.5. Therefore, we have strong evidence to reject $H_0 : \sigma_b^2 = 0$. We can conclude times to occurrence of severe visual loss for the two eyes might be correlated.

4.4.3 Western Kenya Parasitaemia Study

Our third example is from the western Kenya parasitaemia study (McElroy et al., 1997). The study was carried out to investigate the effect of daily mean dose of infective mosquito bite exposure on the risk of parasitaemia, which is an indicator of potential malaria. There were 299 households and 519 children, with multiple children from the same household. The number of the children in each household may be different, so the data are not balanced. The failure time is the time until parasitaemia, and these may be correlated within households. One covariate is quadratic root of daily mean dose of infective mosquito bite (qbite), which might be subject to measurement error, as discussed in Li and Lin (2000). For our purposes, we ignore the possible measurement error and treat this covariate as known. Other covariates include gender, coded as 1 for female and 0 for male; age; and logarithm (base 10) baseline parasitaemia density (denoted as logbase in the model). Because of the possible correlation of the time until parasitaemia, we fitted our proposed subject-specific AFT model as

$$\log T_{ij} = \beta_1 \text{qbite}_{ij} + \beta_2 \text{age}_{ij} + \beta_3 \text{gender}_{ij} + \beta_4 \text{logbase}_{ij} + b_i + e_{ij}.$$
The preferred model takes $K = 2$ with the normal “kernel” based on the HQ information criterion, suggesting that the conditional failure time distribution departs from the lognormal.

Table 4.20 shows the MLEs, estimated standard errors, and 95% Wald confidence intervals. The confidence interval for the bite does not cover zero, so we have strong evidence to support that the time until parasitaemia is about $e^{0.192} - 1 = 0.212$ times shorter if the quadratic root of daily mean dose of infective mosquito bite increases by one unit.

Table 4.20: Inference results for the Kenya parasitaemia dataset. Table entries are the same as in Table 4.18.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>Est. SE</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>qbite</td>
<td>-0.192</td>
<td>0.057</td>
<td>(-0.304, -0.080)</td>
</tr>
<tr>
<td>age</td>
<td>-0.011</td>
<td>0.015</td>
<td>(-0.040, 0.018)</td>
</tr>
<tr>
<td>gender</td>
<td>-0.030</td>
<td>0.047</td>
<td>(-0.122, 0.062)</td>
</tr>
<tr>
<td>logbase</td>
<td>-0.040</td>
<td>0.022</td>
<td>(-0.083, 0.003)</td>
</tr>
<tr>
<td>$\sigma_b$</td>
<td>0.135</td>
<td>0.054</td>
<td>(0.029, 0.241)</td>
</tr>
<tr>
<td>$E(e_{ij})$</td>
<td>3.827</td>
<td>0.105</td>
<td>(3.621, 4.033)</td>
</tr>
<tr>
<td>$sd(e_{ij})$</td>
<td>0.543</td>
<td>0.027</td>
<td>(0.490, 0.596)</td>
</tr>
</tbody>
</table>

We are also interested in whether or not the times until parasitaemia in each household are correlated or not. We carried out the hypothesis test $H_0 : \sigma_b^2 = 0$ vs. $H_a : \sigma_b^2 > 0$. The LRT test statistic is equal to 0.604, which is less than 95% critical point of a mixture of $\chi^2(0)$ and $\chi^2(1)$ with equal probability 0.5. Therefore, there is not enough evidence to reject $H_0 : \sigma_b^2 = 0$. We do not have enough evidence to conclude that times until parasitaemia in each household are correlated.
4.5 Discussion

Our proposed model does not require a parametric assumption on the within-cluster error term. The method is straightforward to implement using MCEM. Because the likelihood is “parametric” for any \( K \) and “kernel,” arbitrary censoring patterns (interval censoring, left truncation etc.) are easily handled. Relaxation of the assumption of normal random effects needs further investigation. Although we do not pursue it here, we note that a similar formulation could be developed for the proportional hazards model with frailty.
Chapter 5

Bivariate Survival Function Estimation

5.1 Introduction

In this chapter, we develop our proposed methods for estimating a bivariate survival function under mild “smoothness” conditions. Our proposed methods do not require the margins of the bivariate function to be the same, but may be simplified to allow common margins in cases where the assumption of common margin is appropriate.

We consider only right-censored data, but remark that, because the SNP approach yeilds a “parametric” likelihood for fixed $K$ and “kernel,” arbitrary censoring patterns are readily handled in principle.
5.2 Notation

Let \( T_i = (T_{i1}, T_{i2})^T \) be a pair of times-to-event for subject \( i \), and let \( C_i = (C_{i1}, C_{i2})^T \) be the corresponding right censoring times, assumed independent of \( T_i \). We do not insist on common margins for \( T_{i1} \) and \( T_{i2} \), so that \( T_{i1} \) and \( T_{i2} \) may represent different phenomena. Let \( \tilde{\Delta}_{ij} = I(T_{ij} \leq C_{ij}) \) and \( V_{ij} = \min(T_{ij}, C_{ij}) \), \( i = 1, \ldots, n, \ j = 1, 2 \). The observed data are i.i.d. pairs \( (V_i, \tilde{\Delta}_i) \), \( V_i = (V_{i1}, V_{i2})^T \), \( \tilde{\Delta}_i = (\tilde{\Delta}_{i1}, \tilde{\Delta}_{i2})^T \), and let \( V = (V_1, \ldots, V_n)^T \), \( \tilde{\Delta} = (\tilde{\Delta}_1, \ldots, \tilde{\Delta}_n)^T \). We assume that \( T_i \) has bivariate density and survival functions \( f \) and \( S \), and that the density satisfies mild “smoothness” conditions as in Gallant and Nychka (1987). We wish to estimate \( S(t_1, t_2) \).

5.3 SNP Density Representation

5.3.1 SNP Representation for Noncommon Margins

If we do not insist on common margins, the potential to incorporate the normal “kernel” suggests the representation of joint failure times as

\[
\begin{pmatrix}
\log T_{i1} \\
\log T_{i2}
\end{pmatrix} =
\begin{pmatrix}
\mu_1 \\
\mu_2
\end{pmatrix} +
\begin{pmatrix}
\sigma_{11} & \sigma_{12} \\
0 & \sigma_{22}
\end{pmatrix}
\begin{pmatrix}
Z_{i1} \\
Z_{i2}
\end{pmatrix},
\]

where \( Z_i = (Z_{i1}, Z_{i2})^T \) has density \( h(z) \in \mathcal{H} \), and \( \mathcal{H} \) is the class of “smooth” densities, so that \( f \in \mathcal{H} \). Following Gallant and Nychka (1987), we represent \( h(z) \) by the general bivariate SNP density \( h_K(z) \) for fixed \( K \), where \( h_K(z) = P_K^2(z)\xi_1(z_1)\xi_1(z_2) \), and \( \xi_1(z) \) is taken to be the standard \( N(0, 1) \) density.

The shape of SNP density is rich enough to represent a wide class of distributions.
Densities with this two dimensional SNP representation might be multi-modal, fat- or thin-tailed relative to the bivariate normal distribution or skewed. Here, we use two examples to show how broad the class of distributions is that the two dimensional SNP density can represent.

One example is with $q = 2$ and $K = 1$, $h_1(z) = P_1(z; a) \varphi(z_1; 0, 1) \varphi(z_2; 0, 1)$, where $P_1(z; a) = a_1 + a_3 z_1 + a_2 z_2$. With $\psi = (1.176, 1.280)^T$, we obtain $a = (0.923, 0.368, 0.110)^T$. Let $\mu_1 = -1.541$, $\mu_2 = -2.730$, $\sigma_1 = -0.071$, $\sigma_2 = 0.127$, and $\sigma_{12} = 0.503$, then based on the SNP representation in (5.1), the joint density of $(\log T_{i1}, \log T_{i2})$ is shown in Figure 5.1.

The other example is for $q = 2$ and $K = 2$, $h_2(z) = P_2(z; a) \varphi(z_1; 0, 1) \varphi(z_2; 0, 1)$, where $P_2(z; a) = a_1 + a_2 z_2 + a_3 z_2^2 + a_4 z_1 + a_5 z_1^2 + a_6 z_1 z_2$.

With $\psi = (0.401, 1.184, 0.647, 1.167, -0.236)^T$, we obtain $a = (0.612, -0.853, -0.147, -0.254, 0.025, -0.075)^T$. Let $\mu_1 = -0.155$, $\mu_2 = -0.539$, $\sigma_1 = 0.020$, $\sigma_2 = 0.333$, and $\sigma_{12} = 0.448$, then Figure 5.2 shows the joint density of $(\log T_{i1}, \log T_{i2})$ based on the SNP representation in (5.1).
Figure 5.1: SNP Density Representation with Normal “Kernel” and $K = 1$
Figure 5.2: SNP Density Representation with Normal “Kernel” and $K = 2$
The alternative potential to incorporate exponential “kernel” suggests the representation of joint failure times as

\[
\begin{pmatrix}
  \log T_{i1} \\
  \log T_{i2}
\end{pmatrix}
= \begin{pmatrix}
  \mu_1 \\
  \mu_2
\end{pmatrix} + \begin{pmatrix}
  e^{\sigma_1} & \sigma_{12} \\
  0 & e^{\sigma_2}
\end{pmatrix}
\begin{pmatrix}
  \log Z_{i1} \\
  \log Z_{i2}
\end{pmatrix},
\]

(5.2)

where \( Z_i = (Z_{i1}, Z_{i2})^T \) has density \( h(z) \in \mathcal{H} \), where \( H \) is the class of “smooth” densities, so that \( f \in \mathcal{H} \). Following Gallant and Nychka (1987), we represent \( h(z) \) by the general bivariate SNP density \( h_K(z) \) for fixed \( K \), where \( h_K(z) = P_1^2(z; a)e^{-z_1}e^{-z_2} \), where \( \xi_1(z) \) is taken to be the standard exponential density \( e^{-z} \).

Again, we use one example to show how well two dimensional SNP density can represent for exponential “kernel.” For \( q = 2 \) and \( K = 1 \), \( h_1(z) = P_1^2(z; a)e^{-z_1}e^{-z_2} \), where \( P_1(z; a) = a_1 + a_3z_1 + a_2z_2 \). With \( \psi = (0.876, 0.063)^T \), we obtain \( a = (0.089, 0.041, 0.639)^T \). Let \( \mu_1 = -2.069, \mu_2 = -0.056, \sigma_1 = 0.586, \sigma_2 = 0.055 \), and \( \sigma_{12} = -0.046 \), then based on the SNP representation in (5.2), the joint density of \((T_{i1}, T_{i2})\) is shown in Figure 5.3.
Figure 5.3: SNP Density Representation with Exponential “Kernel” and $K = 1$
In either case, following the discussion in Chapter 2, in order to avoid numerical unstablity of estimating the coefficients \( a \), we make a polar coordinate transformation to \( \psi \). Letting \( \theta = (\mu_1, \mu_2, \sigma_1, \sigma_2, \sigma_{12}, \psi)^T \), estimation of the bivariate density and survival function corresponds to estimating \( \theta \) for different \( K \) and “kernel.”

### 5.3.2 SNP Representation for Common Margins

In some situations, e.g., studies of organs such as the eyes and kidneys, we may wish to assume the margins for both eyes or kidneys are same. Our proposed method may be simplified to allow for common margins. Under the normal “kernel,” we represent the joint failure times as

\[
\begin{pmatrix}
\log T_{i1} \\
\log T_{i2}
\end{pmatrix}
= \begin{pmatrix}
\mu \\
\mu
\end{pmatrix} + \begin{pmatrix}
e^\sigma & \sigma_{12} \\
\sigma_{12} & e^\sigma
\end{pmatrix}
\begin{pmatrix}
Z_{i1} \\
Z_{i2}
\end{pmatrix},
\]  \hspace{1cm} (5.3)

where \( Z_i = (Z_{i1}, Z_{i2})^T \) has density \( h(z) \in \mathcal{H} \), where \( \mathcal{H} \) is the class of “smooth” densities, so that \( f \in \mathcal{H} \), and represent \( h(z) \) by the general bivariate SNP density \( h_K(z) \) for fixed \( K \), where \( h_K(z) = P_K^2(z) \xi_1(z_1) \xi_1(z_2) \), where \( \xi_1(z) \) is the standard \( N(0,1) \) density.

Similarly, for the exponential “kernel,” we have

\[
\begin{pmatrix}
\log T_{i1} \\
\log T_{i2}
\end{pmatrix}
= \begin{pmatrix}
\mu \\
\mu
\end{pmatrix} + \begin{pmatrix}
e^\sigma & \sigma_{12} \\
\sigma_{12} & e^\sigma
\end{pmatrix}
\begin{pmatrix}
\log Z_{i1} \\
\log Z_{i2}
\end{pmatrix},
\]  \hspace{1cm} (5.4)

where \( Z_i = (Z_{i1}, Z_{i2})^T \) has density \( h(z) \in \mathcal{H} \), where \( \mathcal{H} \) is the class of “smooth” densities, so that \( f \in \mathcal{H} \), and represent \( h(z) \) by the general bivariate SNP density \( h_K(z) \) for fixed \( K \), where \( h_K(z) = P_K^2(z) \xi_1(z_1) \xi_1(z_2) \) and \( \xi_1(z) = e^{-z} \).
To achieve common margins, we add the restriction that each coefficient $a_\lambda$ is invariant with respect to permutation of the index $\lambda$, e.g. for $q = 2$ and $K = 2$, $P_K(z) = a_1 + a_2(z_1 + z_2) + a_3(z_1^2 + z_2^2) + a_4z_1z_2$. Under this formulation, we can easily show that the marginal densities of $Z_{i1}$ and $Z_{i2}$ are same, and given by $a_2(a_1 + a_2z + a_3z^2)\xi_1(z)$. Hence, from (5.3) and (5.4), the marginal densities of $T_{i1}$ and $T_{i2}$ are same. Following the same argument for the reparameterization of coefficients $a$ to $\psi$, here we can define vector $V = (1, U_1 + U_2, U_1^2 + U_2^2, U_1U_2)^T$, where $U_1, U_2 \sim \xi_1(u)$ independently (see Chapter 2), calculate the matrix $A = E(VV^T)$ and follow the same argument to derive the $\psi$. Thus with exponential “kernel,”

$$A = E(VV^T) = \begin{bmatrix}
1 & 2 & 4 & 1 \\
2 & 6 & 16 & 4 \\
4 & 16 & 56 & 12 \\
1 & 4 & 12 & 4
\end{bmatrix},$$

$$B = \begin{bmatrix}
1 & 2 & 4 & 1 \\
0 & \sqrt{2} & 4\sqrt{2} & \sqrt{2} \\
0 & 0 & 2\sqrt{2} & 0 \\
0 & 0 & 0 & 1
\end{bmatrix},$$

$$B^{-1} = \begin{bmatrix}
1 & -\sqrt{2} & \sqrt{2} & 1 \\
0 & \sqrt{2}/2 & -\sqrt{2} & -1 \\
0 & 0 & \sqrt{2}/4 & 0 \\
0 & 0 & 0 & 1
\end{bmatrix}.$$
and
\[ c(\psi) = \begin{bmatrix}
\sin \psi_1 \\
\cos \psi_1 \sin \psi_2 \\
\cos \psi_1 \cos \psi_2 \sin \psi_3 \\
\cos \psi_1 \cos \psi_2 \cos \psi_3 
\end{bmatrix}. \]

Therefore, the coefficients \( a \) in \( P_2(z) \) are given by
\[ a = \begin{bmatrix}
\sin \psi_1 - \sqrt{2} \cos \psi_1 \sin \psi_2 + \sqrt{2} \cos \psi_1 \cos \psi_2 \sin \psi_3 + \cos \psi_1 \cos \psi_2 \cos \psi_3 \\
\sqrt{2}/2 \cos \psi_1 \sin \psi_2 - \sqrt{2} \cos \psi_1 \cos \psi_2 \sin \psi_3 - \cos \psi_1 \cos \psi_2 \cos \psi_3 \\
\sqrt{2}/4 \cos \psi_1 \cos \psi_2 \sin \psi_3 \\
\cos \psi_1 \cos \psi_2 \cos \psi_3 
\end{bmatrix}. \]

Letting \( \theta = (\mu, \sigma, \sigma_{12}, \psi^T)^T \), \( \theta \) then defines the joint density representation.

## 5.4 Inference

### 5.4.1 Likelihood Function

For fixed \( K \) and “kernel,” under any of the models given in (5.1) – (5.4), the contribution to the loglikelihood for \( \theta \) for cluster \( i \) based on data \((V_i, \Delta_i)\) is
\[
\ell(\theta; V_i, \Delta_i) = \sum_{i=1}^{n} \left[ \Delta_{i1} \Delta_{i2} \log \{ f(V_i; \theta) \} + \Delta_{i1} \Delta_{i2} \log \left\{ - \frac{\partial S(V_i; \theta)}{\partial V_{i1}} \right\} \right] + \left( 1 - \Delta_{i1} \right) \Delta_{i2} \log \left\{ - \frac{\partial S(V_i; \theta)}{\partial V_{i2}} \right\} + \left( 1 - \Delta_{i1} \right) \left( 1 - \Delta_{i2} \right) \log \{ S(V_i; \theta) \}. 
\]

Therefore, the loglikelihood for \( \theta \) is
\[
\ell(\theta; V, \Delta) = \sum_{i=1}^{n} \ell(\theta; V_i, \Delta_i). \tag{5.5}
\]
Inference based on the joint density and survival function may be based on maximizing the loglikelihood (5.5). In our implementations, we have used Intel Fortran 95 IMSL library DUMPOL subroutine to maximize the loglikelihood under the Microsoft Visual Studio.net IDE.

**Noncommon Margins**

If the margins of $T_{i1}$ and $T_{i2}$ are not imposed to be common, we may represent $f$ for fixed $K$ and normal “kernel” by

$$f(t_1, t_2) = (e^{\sigma_1} e^{\sigma_2} t_1 t_2)^{-1} P_K^2(z_1, z_2) \xi_1(z_1) \xi_1(z_2),$$

where $z_2 = \frac{\log t_2 - \mu_2}{e^{\sigma_2}},$ $z_1 = \frac{\log t_1 - \mu_1}{e^{\sigma_1}} - \frac{\sigma_{12}}{e^{\sigma_1}} z_2,$ and $\xi_1(z)$ is $N(0, 1)$ density. The joint survival function $S$ is then given by

$$S(t_1, t_2) = P(T_{i1} \geq t_1, T_{i2} \geq t_2) = P(Z_{i2} \geq \frac{\log t_2 - \mu_2}{e^{\sigma_2}}, Z_{i1} \geq \frac{\log t_1 - \mu_1}{e^{\sigma_1}} - \frac{\sigma_{12}}{e^{\sigma_1}} Z_{i2})$$

$$= \int_{r_2}^{\infty} \int_{r_1}^{\infty} P_K^2(z_1, z_2) \xi_1(z_1) \xi_1(z_2) dz_1 dz_2,$$

i.e.,

$$S(t_1, t_2) = \int_{r_2}^{\infty} \int_{r_1}^{\infty} P_K^2(z_1, z_2) \xi_1(z_1) \xi_1(z_2) dz_1 dz_2, \quad (5.6)$$

where $r_1 = \frac{\log t_1 - \mu_1}{e^{\sigma_1}}$ and $r_2 = \frac{\log t_2 - \mu_2}{e^{\sigma_2}}.$

Similarly we may represent by $f$ for fixed $K$ and exponential “kernel” by

$$f(t_1, t_2) = (e^{\sigma_1} e^{\sigma_2} t_1 t_2)^{-1} e^{z_1} e^{z_2} P_K^2(e^{z_1}, e^{z_2}) \xi_1(e^{z_1}) \xi_1(e^{z_2}),$$
where \( z_2 = \frac{\log t_2 - \mu_2}{\sigma_2} \), \( z_1 = \frac{\log t_1 - \mu_1}{\sigma_1} - \frac{\sigma_{12}}{\sigma_1} z_2 \), and \( \xi_1(z) \) is the standard exponential density \( e^{-z} \). The joint survival function \( S \) is then given by

\[
S(t_1, t_2) = P(T_{i1} \geq t_1, T_{i2} \geq t_2)
= P(\log Z_{i2} \geq \frac{\log t_2 - \mu_2}{\sigma_2}, \log Z_{i1} \geq \frac{\log t_1 - \mu_1}{\sigma_1} - \frac{\sigma_{12}}{\sigma_1} \log Z_{i2})
= \int_{r_2}^{\infty} \int_{r_1 - \frac{\sigma_{12}}{\exp(\sigma_1) \sigma_2} y_2}^{\infty} P^2_K(e^{y_1}, e^{y_2}) \xi_1(e^{y_1}) \xi_1(e^{y_2}) e^{y_1 + y_2} dy_1 dy_2,
\]

i.e.,

\[
S(t_1, t_2) = \int_{r_2}^{\infty} \int_{r_1 - \frac{\sigma_{12}}{\exp(\sigma_1) \sigma_2} y_2}^{\infty} P^2_K(e^{y_1}, e^{y_2}) \xi_1(e^{y_1}) \xi_1(e^{y_2}) e^{y_1 + y_2} dy_1 dy_2,
\]

(5.7)

where \( r_1 = \frac{\log t_1 - \mu_1}{\sigma_1} \), \( r_2 = \frac{\log t_2 - \mu_2}{\sigma_2} \), and \( \xi_1(z) \) is the standard exponential density \( e^{-z} \).

The two dimensional integration in (5.6) and (5.7) can be reduced to a one dimension integration. Also

\[
-\frac{\partial}{\partial t_1} \{S(t_1, t_2; \theta)\}_{u,v} = \int_{v}^{\infty} f(u, t_2, \theta) dt_2,
\]

and

\[
-\frac{\partial}{\partial t_2} \{S(t_1, t_2; \theta)\}_{u,v} = \int_{u}^{\infty} f(t_1, v, \theta) dt_1,
\]

which can be approximated by the Gauss-Laguerre quadrature formulae (Evans, 1993) with 8 quadrature points.

**Common Margins**

If the marginal distributions of \( T_{i1} \) and \( T_{i2} \) are assumed to be common, we may represent \( f \) for fixed \( K \) and normal “kernel” by

\[
f(t_1, t_2) = (t_1 t_2 | \sigma_{12}^2 - e^{2\sigma}|)^{-1} P^2_K(z_1, z_2) \xi_1(z_1) \xi_1(z_2),
\]

106
where \( z_2 = \frac{\sigma_{12} \log t_1 - e^\sigma \log t_2 - (\sigma_{12} - e^\sigma) \mu}{\sigma_{12}^2 - e^{2\sigma}}, \) \( z_1 = \frac{\log t_1 - \mu_1}{e^\sigma} - \frac{\sigma_{12}}{e^\sigma} z_2, \) and \( \xi_1(z) \) is \( N(0, 1) \) density.

Similarly we may represent \( f \) for fixed \( K \) and exponential “kernel” by

\[
f(t_1, t_2) = (t_1 t_2 | \sigma_{12}^2 - e^{2\sigma})^{-1} e^{z_1} e^{z_2} P_K^2(e^{z_1}, e^{z_2}) \xi_1(e^{z_1}) \xi_1(e^{z_2}),
\]

where \( z_2 = \frac{\sigma_{12} \log t_1 - e^\sigma \log t_2 - (\sigma_{12} - e^\sigma) \mu}{\sigma_{12}^2 - e^{2\sigma}}, \) \( z_1 = \frac{\log t_1 - \mu_1}{e^\sigma} - \frac{\sigma_{12}}{e^\sigma} z_2, \) and \( \xi_1(z) \) is the standard exponential density \( e^{-z} \).

Then the joint survival function \( S \) is given by

\[
S(t_1, t_2) = P(T_{i1} \geq t_1, T_{i2} \geq t_2)
= \int_{t_1}^{\infty} \int_{t_2}^{\infty} f(u_1, u_2) du_1, du_2.
\]

The two dimensional integration can be approximated by Gauss-Hermite quadrature formulae (Evans, 1993) with 40 points. Also,

\[-\frac{\partial}{\partial t_1} \{S(t_1, t_2; \theta)\} |_{u,v} = \int_{t_2}^{\infty} f(u, t_2, \theta) dt_2,\]

and

\[-\frac{\partial}{\partial t_2} \{S(t_1, t_2; \theta)\} |_{u,v} = \int_{t_1}^{\infty} f(t_1, v, \theta) dt_1,\]

which can be approximated by Hermite quadrature with 40 quadrature points.

**5.4.2 Choosing \( K \) and the “Kernel”**

The procedure for choosing \( K \) and the “kernel” density that produce the best fit for the data is analogous to that presented in Section 3.5. In particular, we inspect
values of a given information criteria (e.g., AIC, BIC, HQ) for each “kernel” (normal or exponential) and $K = 0, 1, 2$ and choosing that combination minimizing the chosen criteria HQ.

5.4.3 Hypothesis Test for Correlation

For the correlated bivariate failure times data analysis, the two failure times within the cluster might be correlated. It is often of interest to test whether the potential correlation among the survival times in each cluster is non-zero, which suggests that a model treating these as mutually independent would be inappropriate. The correlation $\rho$ between $\log T_{i1}$ and $\log T_{i2}$ is defined as

$$\rho = \frac{\text{Cov}(\log T_{i1}, \log T_{i2})}{\sqrt{\text{Var}(\log T_{i1})}\sqrt{\text{Var}(\log T_{i2})}},$$

where

$$\text{Cov}(\log T_{i1}, \log T_{i2}) = \int_0^{\infty} \int_0^{\infty} \{\log t_1 - \text{E}(\log T_{i1})\}\{\log t_2 - \text{E}(\log T_{i2})\} f(t_1, t_2) dt_1 dt_2,$$

$$\text{E}(\log T_{i1}) = \int_0^{\infty} \int_0^{\infty} \log t_1 f(t_1, t_2) dt_1 dt_2,$$

and similarly for $\text{E}(\log T_{i2})$, $\text{Var}(\log T_{i1})$, and $\text{Var}(\log T_{i2})$. Therefore, the hypothesis test for correlation is

$$H_0 : \rho = 0 \ vs. \ H_a : \rho \neq 0. \quad (5.8)$$

We advocate Wald inference for (5.8); specifically, one may construct a test statistic as the estimated correlation by the estimated standard error calculated as described in Section 5.4.4.
5.4.4 Standard Errors and Confidence Intervals

The procedure for calculating the estimated standard errors and confidence intervals is analogous to that discussed in Section 3.6. In particular, The Delta method may be used to calculate the standard error of the expectation, variance of the failure times $T_{ij}$, and the correlation between $T_{i1}$ and $T_{i2}$. We also investigated the performance of a nonparametric bootstrap as an alternative to obtaining standard error for one simulation scenario.
Chapter 6

Simulations and Examples for Bivariate Survival Function Estimation

We conducted several simulation studies with right-censored bivariate survival data to investigate the performance of our proposed methods. The joint failure times were drawn from a variety of bivariate distributions. Generation of random variables from a bivariate distribution is not generally supported in most software packages; thus in Section 6.1, we introduce the algorithm we used to generate bivariate random variables from the bivariate exponential distribution and the bivariate Weibull distribution. In Section 6.2, we show the simulation results for maximum likelihood estimation, including performance of estimated standard errors, coverage probability of Wald 95% confidence intervals, density estimation, and point estimation of the survival function. In Section 6.3, we address hypothesis testing for the correlation discussed in Section 5.4.3. We apply our proposed methods to estimation of the bivariate survival function for the DRS dataset in Section 6.4.
6.1 Bivariate Random Variables Generation

Generation of random samples from a bivariate normal distribution and from a mixture of bivariate normal distributions is straightforward; for example, random samples can be generated from these bivariate distributions using Fortran subroutine RNMVN. However, generation of random samples from other plausible bivariate distributions is not generally supported. Here, we describe the algorithm we use in the simulations to generate bivariate exponential distribution and bivariate Weibull distribution.

6.1.1 Generating Bivariate Exponential Random Variables

In general, the marginal distributions of a bivariate distribution can not fully specify the corresponding bivariate distribution. Frechet (1951) proved that, for given marginal distributions, there are infinitely many corresponding bivariate distribution with these margins (Gumbel, 1980). In Section 1.3.1, we introduced the explicit definition of a bivariate distribution function.

There are two very different approaches to construct bivariate exponential distribution. One is the analytic method, which defines a bivariate density or distribution function explicitly such as Nagao and Kadoya (1971) and Gumbel (1980). Another approach is the empirical method, which constructs two random variables to have the correct margins, then constructs bivariate distribution function through a copula such as Moran (1967) and Marshall and Olkin (1967).

Gumbel (1980) derived bivariate exponential distribution of \((T_1, T_2)\) by explicitly
defining its joint distribution function as

\[ F(t_1, t_2) = 1 - e^{-t_1} - e^{-t_2} + e^{-t_1-t_2-at_1t_2}, \quad t_1, t_2 > 0, \]

where \( a \in [0, 1] \) is the parameter. Then he derived the joint density as

\[ f(t_1, t_2) = e^{-t_1-t_2-at_1t_2}(1 + at_1)(1 + at_2) - a, \quad t_1, t_2 > 0. \] (6.1)

Note, when the parameter \( a = 0 \), following the joint density (6.1), \( T_1 \) and \( T_2 \) are independent, hence the correlation between \( T_1 \) and \( T_2 \) is zero. Devroye (1986) provided several algorithms to generate random variables from a bivariate exponential distribution and a bivariate Weibull distribution. One such algorithm that generates a random sample from a bivariate exponential distribution with parameter \( a \), which comes from Gumbel’s exponential family, is as follows:

1. Generate i.i.d. standard exponential random variates \( X_1, X_2 \).

2. Generate a Uniform(0, 1) random variate \( U \).

3. If \( U \leq \frac{a}{1 + aX_1} \) then

   (a) Generate a standard exponential random variate \( E \).

   (b) Replace \( X_2 \) by \( X_2 + E \).

4. Return \( \left(X_1, \frac{X_2}{1 + aX_1}\right) \).

Letting \( T_1 = X_1 \) and \( T_2 = \frac{X_2}{1 + aX_1} \), then \( (T_1, T_2)^T \) is a random bivariate sample from the bivariate exponential distribution with parameter \( a \).
6.1.2 Generating Bivariate Weibull Random Variables

If a random variable \( X \) has a standard exponential distribution \( \text{Exp}(1) \), then \( X^\frac{1}{\alpha} \) has a Weibull distribution with shape parameter \( \alpha > 0 \) and scale parameter 1, i.e., Weibull(\( \alpha, 1 \)). Thus we can easily generate bivariate Weibull random variables based on the algorithm introduced above for bivariate exponential random variables.

6.2 Maximum Likelihood Estimation

6.2.1 Noncommon Margins

We simulated several scenarios where the data are subject to right-censoring. Under each scenario, we generated 100 Monte Carlo data sets. The censoring rate for each margin is about 30%. The correlation between \( T_{i1} \) and \( T_{i2} \), which is denoted as \( \rho \) in Chapter 5, and value of the survival function \( S(t, t) = P_r(T_{i1} > t, T_{i2} > t) \) are our most interesting parameters, where \( t \) was chosen so that the true value of joint survival function is equal to 0.1, 0.2, 0.3, \ldots, 0.9.

The simulation scenarios are as follows:

1. Generate \((\log T_{i1}, \log T_{i2})^T\) from a bivariate normal distribution with mean \((-1, -2)^T\) and variance-covariance matrix \[
\begin{pmatrix}
1.0 & 0.3 \\
0.3 & 1.0
\end{pmatrix};
\]
the true correlation between \( \log T_{i1} \) and \( \log T_{i2} \) is 0.3; generate \((\log C_{i1}, \log C_{i2})^T\) independently from a bivariate normal distribution with mean \((-0.25, -1.25)^T\) and variance-covariance matrix \[
\begin{pmatrix}
1.0 & 0.3 \\
0.3 & 1.0
\end{pmatrix}; i = 1, \ldots, 300.
2. Generate \((\log T_{i1}, \log T_{i2})^T\) from a mixture of the bivariate normal distribution with mean \((-1, -2)^T\) and variance-covariance matrix as 
\[
\begin{pmatrix}
1.0 & 0.3 \\
0.3 & 1.0
\end{pmatrix}
\]
with probability 0.7 and the bivariate normal distribution with mean \((1, 2)^T\) and variance-covariance matrix 
\[
\begin{pmatrix}
1.0 & 0.5 \\
0.5 & 1.0
\end{pmatrix}
\]
with probability 0.3; the true correlation between \(\log T_{i1}\) and \(\log T_{i2}\) is 0.7; generate \((\log C_{i1}, \log C_{i2})^T\) independently from a bivariate normal distribution with mean \((0.45, 0.25)^T\) and variance-covariance matrix 
\[
\begin{pmatrix}
1.0 & 0.3 \\
0.3 & 1.0
\end{pmatrix}
\]
; \(i = 1, \ldots, 600\).

3. Generate \((T_{i1}, T_{i2})^T\) using Gumbel’s (1980) bivariate exponential distribution algorithm with parameter \(a = 0.3\) and standard exponential margins; the true correlation between \(\log T_{i1}\) and \(\log T_{i2}\) is -0.158; generate independently \(C_{i1} \sim \text{Uniform}(0, 3.2)\) and \(C_{i2} \sim \text{Uniform}(0, 3.2); i = 1, \ldots, 300\).

4. Generate \((T_{i1}, T_{i2})^T\) using Gumbel’s (1980) bivariate exponential distribution algorithm with parameter \(a = 0.3\) and standard exponential margins, and obtain \((T_{1\frac{1}{3}}, T_{2\frac{1}{3}})^T\) from a bivariate Weibull distribution with margins Weibull\((3, 1)\) and Weibull\((2, 1)\); the true correlation between \(\log T_{i1}\) and \(\log T_{i2}\) is -0.162; generate independently \(C_{i1} \sim \text{Uniform}(0, 3.0)\) and \(C_{i2} \sim \text{Uniform}(0, 2.9); i = 1, \ldots, 600\).

We fitted each of models (5.1) and (5.2) for \(K = 0, 1, 2\). We summarize the simulation results in Table 6.1. The table shows the number of data sets out of 100 for which each \(K\) and “kernel” combination was chosen by the HQ criterion. Not surprisingly, for the bivariate normal distribution, the most frequently chosen representation is the normal “kernel” with \(K = 0\). For the bivariate mixture normal
distribution, for most datasets, the normal “kernel” with \( K > 0 \) was chosen. When the true distribution departs from bivariate normality, \( K = 0 \) is not adequate, leading to larger \( K \). For the bivariate exponential distribution and bivariate Weibull distribution, for most datasets, the exponential “kernel” with \( K = 0 \) was chosen, as expected. The approach we propose for detecting and representing departure from normal and exponential is encouraging.

We summarize the MLEs, the average of estimated standard errors, empirical Monte Carlo standard deviation, and Monte Carlo coverage probabilities of 95\% Wald confidence intervals for all scenarios in Tables 6.2–6.6.

For bivariate normal data, all the parameter estimators including those of the survival functions are virtually unbiased. The estimated standard error is very close to the Monte Carlo standard deviation. Most of the coverage probabilities are close to 95\%. The true bivariate normal density and the average of estimated densities chosen by HQ are shown in Figure 6.1. The average of estimate density is very close to true density, showing that the SNP density representation is capable of approximating the underlying true bivariate normal density well.

For the bivariate mixture of normal data, we used a larger sample size with \( n = 600 \) clusters. All the parameter estimators are almost unbiased. However, the standard error estimates for some parameters, such as \( \text{sd}(\log T_{i2}) \), and some survival functions at some time points, are not close to the empirical standard deviation, which lead to coverage probabilities that are too low. We thus investigated the performance of a nonparametric bootstrap as an alternative to obtaining standard errors. For each
dataset, we used 50 bootstrap samples and took $n = 300$ to estimate the standard error. The simulation results are listed in Table 6.4. Many of the coverage probabilities improved, although the coverage probability for the survival function at time point $t_4$ is below 90%. Unfortunately, simulation of the bootstrap is very time consuming, and we only used 50 bootstrap samples. We believe if we were able to increase the bootstrap sample size, the results would improve.

The true bivariate mixture normal density and the average of estimated densities chosen by HQ are shown in Figure 6.2. The average of estimate density is very close to true density, again demonstrating that SNP density representation is capable of approximating the true bivariate mixture normal density well.

For the bivariate exponential data, all the parameter estimators including survival function are again almost unbiased, and the estimated standard error is very close to the empirical standard deviation. Most of the coverage probabilities are around 95%. Only at the early time point $t_1$ is the coverage probability low. The true bivariate exponential density and the average of estimated densities chosen by HQ are shown in Figure 6.3. Again, the average of estimate density is very close to true density. Results for the bivariate Weibull data are similar.
Table 6.1: Number of data sets choosing each $K$-base and “Kernel” combination for noncommon margins

<table>
<thead>
<tr>
<th>True Distribution</th>
<th>Normal Kernel</th>
<th>Exp Kernel</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$K$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0 1 2</td>
<td>0 1 2</td>
</tr>
<tr>
<td>Bivariate Lognormal</td>
<td>94 5 1</td>
<td>0 0 0</td>
</tr>
<tr>
<td>Bivariate Logmixture Normal</td>
<td>0 62 38</td>
<td>0 0 0</td>
</tr>
<tr>
<td>Bivariate Exponential</td>
<td>0 0 1</td>
<td>82 8 9</td>
</tr>
<tr>
<td>Bivariate Weibull</td>
<td>0 0 0</td>
<td>75 21 4</td>
</tr>
</tbody>
</table>

True Distribution is the distribution of $T_{ij}$; $\xi_1 \sim N(0,1)$ means normal “kernel”; $\xi_1 \sim \text{Exp}(1)$ means exponential “kernel”; $K$ is the tuning parameter, which can be 0, 1, 2.

Table 6.2: NonCommon margins bivariate normal simulation results

<table>
<thead>
<tr>
<th>Parameter (True)</th>
<th>Ave.</th>
<th>Est. SE</th>
<th>Emp. SD</th>
<th>Cov. Prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>$E(\log T_{11})$ (-1.000)</td>
<td>-0.989</td>
<td>0.064</td>
<td>0.064</td>
<td>0.97</td>
</tr>
<tr>
<td>$E(\log T_{12})$ (-2.000)</td>
<td>-2.002</td>
<td>0.064</td>
<td>0.067</td>
<td>0.94</td>
</tr>
<tr>
<td>sd($\log T_{11}$) (1.000)</td>
<td>1.013</td>
<td>0.053</td>
<td>0.064</td>
<td>0.95</td>
</tr>
<tr>
<td>sd($\log T_{12}$) (1.000)</td>
<td>1.007</td>
<td>0.052</td>
<td>0.068</td>
<td>0.96</td>
</tr>
<tr>
<td>$\rho$ (0.300)</td>
<td>0.310</td>
<td>0.074</td>
<td>0.064</td>
<td>0.97</td>
</tr>
<tr>
<td>$S(t_1, t_1) = 0.9$</td>
<td>0.898</td>
<td>0.014</td>
<td>0.013</td>
<td>0.97</td>
</tr>
<tr>
<td>$S(t_2, t_2) = 0.8$</td>
<td>0.796</td>
<td>0.019</td>
<td>0.018</td>
<td>0.96</td>
</tr>
<tr>
<td>$S(t_3, t_3) = 0.7$</td>
<td>0.697</td>
<td>0.022</td>
<td>0.021</td>
<td>0.95</td>
</tr>
<tr>
<td>$S(t_4, t_4) = 0.6$</td>
<td>0.598</td>
<td>0.023</td>
<td>0.023</td>
<td>0.95</td>
</tr>
<tr>
<td>$S(t_5, t_5) = 0.5$</td>
<td>0.499</td>
<td>0.024</td>
<td>0.025</td>
<td>0.96</td>
</tr>
<tr>
<td>$S(t_6, t_6) = 0.4$</td>
<td>0.398</td>
<td>0.024</td>
<td>0.025</td>
<td>0.95</td>
</tr>
<tr>
<td>$S(t_7, t_7) = 0.3$</td>
<td>0.300</td>
<td>0.024</td>
<td>0.025</td>
<td>0.94</td>
</tr>
<tr>
<td>$S(t_8, t_8) = 0.2$</td>
<td>0.201</td>
<td>0.021</td>
<td>0.022</td>
<td>0.93</td>
</tr>
<tr>
<td>$S(t_9, t_9) = 0.1$</td>
<td>0.102</td>
<td>0.016</td>
<td>0.017</td>
<td>0.93</td>
</tr>
</tbody>
</table>

$S(t_j, t_j) = P(T_{11} \geq t_j, T_{12} \geq t_j)$, where $j = 1, \ldots, 9$; $\rho$ is the correlation between $T_{11}$ and $T_{12}$; Ave. is the average of the estimates over 100 data sets; Estimated SE is the average of the estimated standard errors; Emp. SD is the empirical standard deviation; Cov. Prob. is 95% Wald confidence interval coverage probability.
Figure 6.1: (a) True bivariate normal density (b) Average of 100 density estimates
Table 6.3: Noncommon margins bivariate mixture normal simulation results. Table entries are same as in Table 6.2.

<table>
<thead>
<tr>
<th>Parameter (True)</th>
<th>Ave.</th>
<th>Est. SE</th>
<th>Emp. SD</th>
<th>Cov. Prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>E(log $T_{i1}$) (-0.400)</td>
<td>-0.394</td>
<td>0.062</td>
<td>0.059</td>
<td>0.96</td>
</tr>
<tr>
<td>E(log $T_{i2}$) (-0.800)</td>
<td>-0.765</td>
<td>0.096</td>
<td>0.117</td>
<td>0.91</td>
</tr>
<tr>
<td>sd(log $T_{i1}$) (1.355)</td>
<td>1.369</td>
<td>0.053</td>
<td>0.063</td>
<td>0.94</td>
</tr>
<tr>
<td>sd(log $T_{i2}$) (2.089)</td>
<td>2.139</td>
<td>0.077</td>
<td>0.127</td>
<td>0.80</td>
</tr>
<tr>
<td>$\rho$ (0.700)</td>
<td>0.694</td>
<td>0.025</td>
<td>0.064</td>
<td>0.96</td>
</tr>
<tr>
<td>$S(t_1, t_1) = 0.9$</td>
<td>0.901</td>
<td>0.010</td>
<td>0.009</td>
<td>0.96</td>
</tr>
<tr>
<td>$S(t_2, t_2) = 0.8$</td>
<td>0.806</td>
<td>0.013</td>
<td>0.013</td>
<td>0.92</td>
</tr>
<tr>
<td>$S(t_3, t_3) = 0.7$</td>
<td>0.709</td>
<td>0.016</td>
<td>0.016</td>
<td>0.89</td>
</tr>
<tr>
<td>$S(t_4, t_4) = 0.6$</td>
<td>0.608</td>
<td>0.017</td>
<td>0.018</td>
<td>0.89</td>
</tr>
<tr>
<td>$S(t_5, t_5) = 0.5$</td>
<td>0.502</td>
<td>0.017</td>
<td>0.020</td>
<td>0.89</td>
</tr>
<tr>
<td>$S(t_6, t_6) = 0.4$</td>
<td>0.393</td>
<td>0.018</td>
<td>0.020</td>
<td>0.90</td>
</tr>
<tr>
<td>$S(t_7, t_7) = 0.3$</td>
<td>0.291</td>
<td>0.018</td>
<td>0.021</td>
<td>0.87</td>
</tr>
<tr>
<td>$S(t_8, t_8) = 0.2$</td>
<td>0.202</td>
<td>0.016</td>
<td>0.017</td>
<td>0.94</td>
</tr>
<tr>
<td>$S(t_9, t_9) = 0.1$</td>
<td>0.096</td>
<td>0.013</td>
<td>0.015</td>
<td>0.89</td>
</tr>
</tbody>
</table>

Figure 6.2: (a) True bivariate mixture normal density (b) Average of 100 density estimates
Table 6.4: Noncommon margins bivariate mixture normal simulation results with bootstrap standard error.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$E(\log T_{i1})$ (-0.400)</td>
<td>-0.386</td>
<td>0.088</td>
<td>0.096</td>
<td>0.93</td>
</tr>
<tr>
<td>$E(\log T_{i2})$ (-0.800)</td>
<td>-0.753</td>
<td>0.134</td>
<td>0.140</td>
<td>0.91</td>
</tr>
<tr>
<td>$sd(\log T_{i1})$ (1.355)</td>
<td>1.366</td>
<td>0.078</td>
<td>0.080</td>
<td>0.94</td>
</tr>
<tr>
<td>$sd(\log T_{i2})$ (2.089)</td>
<td>2.130</td>
<td>0.108</td>
<td>0.146</td>
<td>0.92</td>
</tr>
<tr>
<td>$\rho$ (0.700)</td>
<td>0.694</td>
<td>0.037</td>
<td>0.035</td>
<td>0.93</td>
</tr>
<tr>
<td>$S(t_1, t_1) = 0.9$</td>
<td>0.903</td>
<td>0.014</td>
<td>0.014</td>
<td>0.92</td>
</tr>
<tr>
<td>$S(t_2, t_2) = 0.8$</td>
<td>0.810</td>
<td>0.019</td>
<td>0.020</td>
<td>0.91</td>
</tr>
<tr>
<td>$S(t_3, t_3) = 0.7$</td>
<td>0.714</td>
<td>0.022</td>
<td>0.024</td>
<td>0.91</td>
</tr>
<tr>
<td>$S(t_4, t_4) = 0.6$</td>
<td>0.613</td>
<td>0.024</td>
<td>0.027</td>
<td>0.88</td>
</tr>
<tr>
<td>$S(t_5, t_5) = 0.5$</td>
<td>0.507</td>
<td>0.025</td>
<td>0.028</td>
<td>0.92</td>
</tr>
<tr>
<td>$S(t_6, t_6) = 0.4$</td>
<td>0.398</td>
<td>0.025</td>
<td>0.028</td>
<td>0.91</td>
</tr>
<tr>
<td>$S(t_7, t_7) = 0.3$</td>
<td>0.295</td>
<td>0.025</td>
<td>0.028</td>
<td>0.91</td>
</tr>
<tr>
<td>$S(t_8, t_8) = 0.2$</td>
<td>0.204</td>
<td>0.023</td>
<td>0.025</td>
<td>0.95</td>
</tr>
<tr>
<td>$S(t_9, t_9) = 0.1$</td>
<td>0.095</td>
<td>0.018</td>
<td>0.019</td>
<td>0.91</td>
</tr>
</tbody>
</table>

Boot SE. is the nonparametric bootstrap standard error based on 50 bootstrap samples.
Table 6.5: Noncommon margins bivariate exponential simulation results. Table entries are same as in Table 6.2.

<table>
<thead>
<tr>
<th>Parameter (True)</th>
<th>Ave.</th>
<th>Est. SE</th>
<th>Emp. SD</th>
<th>Cov. Prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>$E(\log T_{i1})$ (-0.577)</td>
<td>-0.579</td>
<td>0.077</td>
<td>0.079</td>
<td>0.97</td>
</tr>
<tr>
<td>$E(\log T_{i2})$ (-0.577)</td>
<td>-0.576</td>
<td>0.078</td>
<td>0.090</td>
<td>0.91</td>
</tr>
<tr>
<td>$sd(\log T_{i1})$ (1.283)</td>
<td>1.277</td>
<td>0.072</td>
<td>0.076</td>
<td>0.91</td>
</tr>
<tr>
<td>$sd(\log T_{i2})$ (1.283)</td>
<td>1.296</td>
<td>0.086</td>
<td>0.085</td>
<td>0.93</td>
</tr>
<tr>
<td>$\rho$ (-0.158)</td>
<td>-0.158</td>
<td>0.063</td>
<td>0.067</td>
<td>0.94</td>
</tr>
<tr>
<td>$S(t_1, t_1) = 0.9$</td>
<td>0.901</td>
<td>0.012</td>
<td>0.014</td>
<td>0.87</td>
</tr>
<tr>
<td>$S(t_2, t_2) = 0.8$</td>
<td>0.801</td>
<td>0.018</td>
<td>0.020</td>
<td>0.94</td>
</tr>
<tr>
<td>$S(t_3, t_3) = 0.7$</td>
<td>0.700</td>
<td>0.021</td>
<td>0.023</td>
<td>0.95</td>
</tr>
<tr>
<td>$S(t_4, t_4) = 0.6$</td>
<td>0.598</td>
<td>0.023</td>
<td>0.025</td>
<td>0.95</td>
</tr>
<tr>
<td>$S(t_5, t_5) = 0.5$</td>
<td>0.498</td>
<td>0.024</td>
<td>0.026</td>
<td>0.94</td>
</tr>
<tr>
<td>$S(t_6, t_6) = 0.4$</td>
<td>0.399</td>
<td>0.024</td>
<td>0.025</td>
<td>0.95</td>
</tr>
<tr>
<td>$S(t_7, t_7) = 0.3$</td>
<td>0.301</td>
<td>0.022</td>
<td>0.024</td>
<td>0.94</td>
</tr>
<tr>
<td>$S(t_8, t_8) = 0.2$</td>
<td>0.203</td>
<td>0.020</td>
<td>0.021</td>
<td>0.95</td>
</tr>
<tr>
<td>$S(t_9, t_9) = 0.1$</td>
<td>0.107</td>
<td>0.015</td>
<td>0.017</td>
<td>0.94</td>
</tr>
</tbody>
</table>

Figure 6.3: (a) True bivariate exponential density (b) Average of 100 density estimates
Table 6.6: Noncommon margins bivariate Weibull simulation results. Table entries are same as in Table 6.2.

<table>
<thead>
<tr>
<th>Parameter (True)</th>
<th>Ave.</th>
<th>Est. SE</th>
<th>Emp. SD</th>
<th>Cov. Prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>E(\log T_{i1}) (-0.195)</td>
<td>-0.195</td>
<td>0.019</td>
<td>0.020</td>
<td>0.95</td>
</tr>
<tr>
<td>E(\log T_{i2}) (-0.286)</td>
<td>-0.288</td>
<td>0.027</td>
<td>0.028</td>
<td>0.92</td>
</tr>
<tr>
<td>sd(\log T_{i1}) (0.428)</td>
<td>0.429</td>
<td>0.016</td>
<td>0.018</td>
<td>0.91</td>
</tr>
<tr>
<td>sd(\log T_{i2}) (0.641)</td>
<td>0.642</td>
<td>0.021</td>
<td>0.029</td>
<td>0.86</td>
</tr>
<tr>
<td>\rho (-0.162)</td>
<td>-0.157</td>
<td>0.046</td>
<td>0.051</td>
<td>0.94</td>
</tr>
<tr>
<td>S(t_1, t_1) = 0.9</td>
<td>0.900</td>
<td>0.008</td>
<td>0.010</td>
<td>0.90</td>
</tr>
<tr>
<td>S(t_2, t_2) = 0.8</td>
<td>0.800</td>
<td>0.012</td>
<td>0.014</td>
<td>0.92</td>
</tr>
<tr>
<td>S(t_3, t_3) = 0.7</td>
<td>0.701</td>
<td>0.015</td>
<td>0.016</td>
<td>0.94</td>
</tr>
<tr>
<td>S(t_4, t_4) = 0.6</td>
<td>0.601</td>
<td>0.016</td>
<td>0.017</td>
<td>0.96</td>
</tr>
<tr>
<td>S(t_5, t_5) = 0.5</td>
<td>0.501</td>
<td>0.017</td>
<td>0.017</td>
<td>0.95</td>
</tr>
<tr>
<td>S(t_6, t_6) = 0.4</td>
<td>0.403</td>
<td>0.017</td>
<td>0.017</td>
<td>0.96</td>
</tr>
<tr>
<td>S(t_7, t_7) = 0.3</td>
<td>0.305</td>
<td>0.016</td>
<td>0.016</td>
<td>0.95</td>
</tr>
<tr>
<td>S(t_8, t_8) = 0.2</td>
<td>0.207</td>
<td>0.015</td>
<td>0.014</td>
<td>0.95</td>
</tr>
<tr>
<td>S(t_9, t_9) = 0.1</td>
<td>0.108</td>
<td>0.011</td>
<td>0.011</td>
<td>0.91</td>
</tr>
</tbody>
</table>

Figure 6.4: (a) True bivariate Weibull density (b) Average of 100 density estimates
6.2.2 Common Margins

We simulated several scenarios where the data are subject right-censoring and are drawn from distributions with common margins. Under each scenario, we generated 100 Monte Carlo data sets. The censoring rate for each margin is about 30%. We are interested in the correlation $\rho$ and values of the survival function $S(t, t)$ at various values of $t$ were considered, where $t$ was chosen so that the value of the joint survival function is equal to 0.1, 0.2, 0.3, $\cdots$, 0.9.

The simulation scenarios are as follows:

1. Generate $(\log T_{i1}, \log T_{i2})^T$ from a bivariate normal distribution with mean $(-1, -1)^T$ and variance-covariance matrix

   \[
   \begin{pmatrix}
   1.0 & 0.3 \\
   0.3 & 1.0
   \end{pmatrix};
   \]

   the true correlation between $\log T_{i1}$ and $\log T_{i2}$ is 0.3; generate $(\log C_{i1}, \log C_{i2})^T$ independently from a bivariate normal distribution with mean $(-0.25, -0.25)^T$ and variance-covariance matrix as

   \[
   \begin{pmatrix}
   1.0 & 0.3 \\
   0.3 & 1.0
   \end{pmatrix};
   \]

   $i = 1, \ldots, 300$.

2. Generate $(\log T_{i1}, \log T_{i2})^T$ from a mixture of bivariate normal distribution with mean $(-1, -1)^T$ and variance-covariance matrix

   \[
   \begin{pmatrix}
   1.0 & 0.3 \\
   0.3 & 1.0
   \end{pmatrix}
   \]

   with probability 0.7 and a bivariate normal distribution with mean $(1, 1)^T$ and variance-covariance matrix

   \[
   \begin{pmatrix}
   1.0 & 0.3 \\
   0.3 & 1.0
   \end{pmatrix}
   \]

   with probability 0.3; the true correlation between $\log T_{i1}$ and $\log T_{i2}$ is 0.622; generate $(\log C_{i1}, \log C_{i2})^T$ independently from a bivariate normal distribution with mean $(0.45, 0.45)^T$ and variance-covariance
matrix \[
\begin{pmatrix}
1.0 & 0.3 \\
0.3 & 1.0 \\
\end{pmatrix}; i = 1, \ldots, 600.
\]

3. Generate \((T_{i1}, T_{i2})^T\) from Gumbel’s bivariate exponential distribution with parameter \(a = 0.3\) and standard exponential margins; the true correlation between \(\log T_{i1}\) and \(\log T_{i2}\) is \(-0.158\); generate \(C_{i1} \sim \text{Uniform}(0, 3.2)\), and \(C_{i2} \sim \text{Uniform}(0, 3.2)\); \(i = 1, \ldots, 300\).

4. Generate \((T_{i1}, T_{i2})^T\) from Gumbel’s bivariate exponential distribution with parameter \(a = 0.3\), and then form \((T_{i1}^{\frac{1}{3}}, T_{i2}^{\frac{1}{3}})^T\), which has a bivariate Weibull distribution with margins Weibull\((3, 1)\) and Weibull\((3, 1)\); the true correlation between \(\log T_{i1}\) and \(\log T_{i2}\) is \(-0.163\); generate \(C_{i1} \sim \text{Uniform}(0, 3.0)\) and \(C_{i2} \sim \text{Uniform}(0, 3.0)\); \(i = 1, \ldots, 300\).

We fitted models (5.3) and (5.4) for each of \(K = 0, 1, 2\), and we summarize the simulation results in Table 6.7. The table shows the number of data sets out of 100 for which each \(K\) and “kernel” combination was chosen by the HQ criterion. The results are similar to those for the noncommon margins case.

We summarize the results for all scenarios in Tables 6.8–6.11. True and estimated densities are shown in Figures 6.5 – 6.8. The results are similar to those in the noncommon margins case; however, while some low coverage probabilities were seen in that settings, here, where common margins are imposed, the coverage probabilities are all close to the nominal level of 0.95. Evidently, when the margins are truly the same, there is some benefit to using this more parsimonious representation.
Table 6.7: Number of data sets choosing each $K$-base and “Kernel” combination for common margins

<table>
<thead>
<tr>
<th>True Distribution</th>
<th>Normal Kernel</th>
<th>Exp Kernel</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$K$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0 1 2</td>
<td>0 1 2</td>
</tr>
<tr>
<td>Bivariate Lognormal</td>
<td>94 4 2</td>
<td>0 0 0</td>
</tr>
<tr>
<td>Bivariate Logmixture Normal</td>
<td>0 2 98</td>
<td>0 0 0</td>
</tr>
<tr>
<td>Bivariate Exponential</td>
<td>0 0 0</td>
<td>95 4 1</td>
</tr>
<tr>
<td>Bivariate Weibull</td>
<td>0 0 0</td>
<td>60 32 8</td>
</tr>
</tbody>
</table>

True Distribution is the distribution of $T_{ij}$; $\xi_1 \sim N(0, 1)$ means normal “kernel”; $\xi_1 \sim \text{Exp}(1)$ means exponential “kernel.” $K$ is the tuning parameter, which can be 0, 1, 2.

Table 6.8: Common margins bivariate normal simulation results. Table entries are same as in Table 6.2.

<table>
<thead>
<tr>
<th>Parameter (True)</th>
<th>Ave.</th>
<th>Est. SE</th>
<th>Emp. SD</th>
<th>Cov. Prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>$E(\log T_{11})$ (-1.000)</td>
<td>-1.004</td>
<td>0.050</td>
<td>0.045</td>
<td>0.98</td>
</tr>
<tr>
<td>$sd(\log T_{11})$ (1.000)</td>
<td>1.004</td>
<td>0.050</td>
<td>0.043</td>
<td>0.94</td>
</tr>
<tr>
<td>$\rho$ (0.300)</td>
<td>0.302</td>
<td>0.071</td>
<td>0.071</td>
<td>0.93</td>
</tr>
<tr>
<td>$S(t_1, t_1) = 0.9$</td>
<td>0.896</td>
<td>0.014</td>
<td>0.012</td>
<td>0.96</td>
</tr>
<tr>
<td>$S(t_2, t_2) = 0.8$</td>
<td>0.796</td>
<td>0.019</td>
<td>0.016</td>
<td>0.97</td>
</tr>
<tr>
<td>$S(t_3, t_3) = 0.7$</td>
<td>0.696</td>
<td>0.022</td>
<td>0.018</td>
<td>0.98</td>
</tr>
<tr>
<td>$S(t_4, t_4) = 0.6$</td>
<td>0.598</td>
<td>0.023</td>
<td>0.019</td>
<td>0.99</td>
</tr>
<tr>
<td>$S(t_5, t_5) = 0.5$</td>
<td>0.497</td>
<td>0.024</td>
<td>0.021</td>
<td>0.99</td>
</tr>
<tr>
<td>$S(t_6, t_6) = 0.4$</td>
<td>0.398</td>
<td>0.024</td>
<td>0.021</td>
<td>0.98</td>
</tr>
<tr>
<td>$S(t_7, t_7) = 0.3$</td>
<td>0.299</td>
<td>0.023</td>
<td>0.021</td>
<td>0.98</td>
</tr>
<tr>
<td>$S(t_8, t_8) = 0.2$</td>
<td>0.199</td>
<td>0.020</td>
<td>0.020</td>
<td>0.98</td>
</tr>
<tr>
<td>$S(t_9, t_9) = 0.1$</td>
<td>0.099</td>
<td>0.015</td>
<td>0.015</td>
<td>0.97</td>
</tr>
</tbody>
</table>
Figure 6.5: (a) True bivariate normal density (b) Average of 100 density estimates
Table 6.9: Common margins bivariate mixture of normal simulation results. Table entries are same as in Table 6.2.

<table>
<thead>
<tr>
<th>Parameter (True)</th>
<th>Ave.</th>
<th>Est. SE</th>
<th>Emp. SD</th>
<th>Cov. Prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>E(log $T_{i1}$) (-0.400)</td>
<td>-0.400</td>
<td>0.054</td>
<td>0.050</td>
<td>0.97</td>
</tr>
<tr>
<td>sd(log $T_{i1}$) (1.355)</td>
<td>1.369</td>
<td>0.044</td>
<td>0.045</td>
<td>0.96</td>
</tr>
<tr>
<td>$\rho$ (0.622)</td>
<td>0.627</td>
<td>0.031</td>
<td>0.032</td>
<td>0.93</td>
</tr>
<tr>
<td>$S(t_1, t_1)$</td>
<td>0.899</td>
<td>0.010</td>
<td>0.010</td>
<td>0.96</td>
</tr>
<tr>
<td>$S(t_2, t_2)$</td>
<td>0.797</td>
<td>0.014</td>
<td>0.013</td>
<td>0.97</td>
</tr>
<tr>
<td>$S(t_3, t_3)$</td>
<td>0.698</td>
<td>0.016</td>
<td>0.015</td>
<td>0.98</td>
</tr>
<tr>
<td>$S(t_4, t_4)$</td>
<td>0.599</td>
<td>0.018</td>
<td>0.017</td>
<td>0.94</td>
</tr>
<tr>
<td>$S(t_5, t_5)$</td>
<td>0.500</td>
<td>0.019</td>
<td>0.019</td>
<td>0.93</td>
</tr>
<tr>
<td>$S(t_6, t_6)$</td>
<td>0.400</td>
<td>0.019</td>
<td>0.020</td>
<td>0.93</td>
</tr>
<tr>
<td>$S(t_7, t_7)$</td>
<td>0.299</td>
<td>0.018</td>
<td>0.019</td>
<td>0.92</td>
</tr>
<tr>
<td>$S(t_8, t_8)$</td>
<td>0.199</td>
<td>0.016</td>
<td>0.017</td>
<td>0.94</td>
</tr>
<tr>
<td>$S(t_9, t_9)$</td>
<td>0.102</td>
<td>0.014</td>
<td>0.014</td>
<td>0.96</td>
</tr>
</tbody>
</table>
Figure 6.6: (a) True bivariate mixture normal density (b) Average of 100 density estimates
Table 6.10: Common margins bivariate exponential simulation results. Table entries are same as in Table 6.2.

<table>
<thead>
<tr>
<th>Parameter (True)</th>
<th>Ave.</th>
<th>Est. SE</th>
<th>Emp. SD</th>
<th>Cov. Prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>E(log $T_{i1}$) (-0.577)</td>
<td>-0.582</td>
<td>0.050</td>
<td>0.053</td>
<td>0.93</td>
</tr>
<tr>
<td>sd(log $T_{i1}$) (1.283)</td>
<td>1.281</td>
<td>0.055</td>
<td>0.054</td>
<td>0.98</td>
</tr>
<tr>
<td>$\rho$ (-0.158)</td>
<td>-0.178</td>
<td>0.067</td>
<td>0.066</td>
<td>0.92</td>
</tr>
<tr>
<td>$S(t_1, t_1) = 0.9$</td>
<td>0.899</td>
<td>0.013</td>
<td>0.013</td>
<td>0.94</td>
</tr>
<tr>
<td>$S(t_2, t_2) = 0.8$</td>
<td>0.799</td>
<td>0.019</td>
<td>0.019</td>
<td>0.95</td>
</tr>
<tr>
<td>$S(t_3, t_3) = 0.7$</td>
<td>0.698</td>
<td>0.022</td>
<td>0.023</td>
<td>0.95</td>
</tr>
<tr>
<td>$S(t_4, t_4) = 0.6$</td>
<td>0.596</td>
<td>0.024</td>
<td>0.025</td>
<td>0.96</td>
</tr>
<tr>
<td>$S(t_5, t_5) = 0.5$</td>
<td>0.496</td>
<td>0.025</td>
<td>0.025</td>
<td>0.96</td>
</tr>
<tr>
<td>$S(t_6, t_6) = 0.4$</td>
<td>0.397</td>
<td>0.024</td>
<td>0.025</td>
<td>0.97</td>
</tr>
<tr>
<td>$S(t_7, t_7) = 0.3$</td>
<td>0.299</td>
<td>0.023</td>
<td>0.023</td>
<td>0.94</td>
</tr>
<tr>
<td>$S(t_8, t_8) = 0.2$</td>
<td>0.202</td>
<td>0.020</td>
<td>0.021</td>
<td>0.93</td>
</tr>
<tr>
<td>$S(t_9, t_9) = 0.1$</td>
<td>0.105</td>
<td>0.015</td>
<td>0.016</td>
<td>0.92</td>
</tr>
</tbody>
</table>
Figure 6.7: (a) True bivariate exponential density (b) Average of 100 density estimates
Table 6.11: Common margins bivariate Weibull simulation results. Table entries are same as in Table 6.2.

<table>
<thead>
<tr>
<th>Parameter (True)</th>
<th>Ave.</th>
<th>Est. SE</th>
<th>Emp. SD</th>
<th>Cov. Prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>E(log $T_{i1}$) (-0.193)</td>
<td>-0.193</td>
<td>0.013</td>
<td>0.012</td>
<td>0.94</td>
</tr>
<tr>
<td>sd(log $T_{i1}$) (0.428)</td>
<td>0.429</td>
<td>0.012</td>
<td>0.012</td>
<td>0.93</td>
</tr>
<tr>
<td>$\rho$ (-0.163)</td>
<td>-0.178</td>
<td>0.046</td>
<td>0.047</td>
<td>0.92</td>
</tr>
<tr>
<td>$S(t_1, t_1) = 0.9$</td>
<td>0.898</td>
<td>0.009</td>
<td>0.009</td>
<td>0.94</td>
</tr>
<tr>
<td>$S(t_2, t_2) = 0.8$</td>
<td>0.798</td>
<td>0.013</td>
<td>0.013</td>
<td>0.93</td>
</tr>
<tr>
<td>$S(t_3, t_3) = 0.7$</td>
<td>0.697</td>
<td>0.016</td>
<td>0.016</td>
<td>0.93</td>
</tr>
<tr>
<td>$S(t_4, t_4) = 0.6$</td>
<td>0.596</td>
<td>0.017</td>
<td>0.017</td>
<td>0.94</td>
</tr>
<tr>
<td>$S(t_5, t_5) = 0.5$</td>
<td>0.497</td>
<td>0.018</td>
<td>0.017</td>
<td>0.95</td>
</tr>
<tr>
<td>$S(t_6, t_6) = 0.4$</td>
<td>0.398</td>
<td>0.018</td>
<td>0.017</td>
<td>0.93</td>
</tr>
<tr>
<td>$S(t_7, t_7) = 0.3$</td>
<td>0.300</td>
<td>0.017</td>
<td>0.016</td>
<td>0.93</td>
</tr>
<tr>
<td>$S(t_8, t_8) = 0.2$</td>
<td>0.203</td>
<td>0.015</td>
<td>0.014</td>
<td>0.94</td>
</tr>
<tr>
<td>$S(t_9, t_9) = 0.1$</td>
<td>0.106</td>
<td>0.011</td>
<td>0.011</td>
<td>0.95</td>
</tr>
</tbody>
</table>
Figure 6.8: (a) True bivariate weibull density (b) Average of 100 density estimates
6.3 Hypothesis Test for Correlation

We discussed hypothesis testing for correlation in Section 5.4.3. Here, under noncommon margin scenario, we conducted simulation studies to investigate if the nominal size of Wald test is close to 0.05 by using the level-\( \alpha = 0.05 \) Wald test to test hypothesis

\[
H_0 : \rho = 0 \quad \text{vs.} \quad H_a : \rho \neq 0.
\]

We simulated 100 datasets under the null hypothesis \( \rho = 0 \). The censoring rate for each margin is about 30%.

The simulation scenarios are as follows:

1. Generate \((\log T_{i1}, \log T_{i2})^T\) from a bivariate normal distribution with mean \((-1, -2)^T\) and variance-covariance matrix \(
\begin{pmatrix}
1.0 & 0.0 \\
0.0 & 1.0
\end{pmatrix}
\); generate \((\log C_{i1}, \log C_{i2})^T\) independently from a bivariate normal distribution with mean \((-0.25, -1.25)^T\) and variance-covariance matrix \(
\begin{pmatrix}
1.0 & 0.0 \\
0.0 & 1.0
\end{pmatrix}
\); \(i = 1, \ldots, 600\).

2. Generate \((T_{i1}, T_{i2})^T\) using Gumbel’s bivariate exponential distribution algorithm with parameter \(a = 0.0\) and standard exponential margins; generate independently \(C_{i1} \sim \text{Uniform}(0, 3.2)\) and \(C_{i2} \sim \text{Uniform}(0, 3.2); i = 1, \ldots, 600\).

3. Generate \((T_{i1}, T_{i2})^T\) using Gumbel’s bivariate exponential distribution algorithm with parameter \(a = 0.0\) and standard exponential margins, and obtain \((T_1^{\frac{1}{3}}, T_2^{\frac{1}{3}})^T\) from a bivariate Weibull distribution with margins Weibull(3, 1)
and Weibull(2, 1); generate independently $C_{i1} \sim \text{Uniform}(0, 3.0)$ and $C_{i2} \sim \text{Uniform}(0, 2.9)$; $i = 1, \ldots, 600$.

We fitted each of models (5.1) and (5.2) for $K = 0, 1, 2$. Table 6.12 shows the Monte Carlo rejection rate for each simulation scenario and show that the test appears to respect the nominal level.

Table 6.12: Hypothesis test for correlation

<table>
<thead>
<tr>
<th>True Distribution</th>
<th>$\alpha$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bivariate Lognormal</td>
<td>0.03</td>
</tr>
<tr>
<td>Bivariate Exponential</td>
<td>0.05</td>
</tr>
<tr>
<td>Bivariate Weibull</td>
<td>0.05</td>
</tr>
</tbody>
</table>

6.4 Application

We applied our proposed method to Diabetic Retinopathy Study, which was described in Section 4.4.2. We consider estimation of the joint survival distribution for treated and untreated eyes. Let $T_{i1}$ be the failure time for the treated eye and $T_{i2}$ be the failure time for the untreated eye. We fitted $K = 0, 1, 2$ for both the normal “kernel” and exponential “kernel.” The HQ criterion chose the exponential “kernel” with $K = 2$. Because one eye is treated and the other eye is not treated, it makes sense to assume the margins for two eyes are not same. Table 6.13 shows the MLEs, analytical standard error and 95% Wald confidence intervals based on analytical and bootstrap standard errors (based on 100 bootstrap samples). The correlation between the two failure times for the two eyes is a parameter of interest. Note that the confidence
interval for this parameter do not cover zero, indicating evidence to conclude that there is correlation between the failure times of the treated eye and untreated eye. Also, Table 6.13 shows the estimation of joint survival function at 9 time points given by 20, 30, 40, 50, 60, 70, 80, 150, 360. Figure 6.9 shows the estimated bivariate density. Figure 6.10 shows the estimated marginal density of the failure time of treated eye $T_{i1}$ and the estimated marginal density of the failure time of untreated eye $T_{i2}$. It shows the two margins are not same, further prove our assumption. Figure 6.11 shows the estimated bivariate survival function $Pr(T_{i1} \geq t_1, T_{i2} \geq t_2)$. Figure 6.12 shows the estimated survival function $Pr(T_{i1} \geq t, T_{i2} \geq t)$. 

135
Table 6.13: Inference for DRS data set

<table>
<thead>
<tr>
<th>Parameter</th>
<th>MLE</th>
<th>Est. SE</th>
<th>Boot SE</th>
<th>95% CI*</th>
<th>95% CI**</th>
</tr>
</thead>
<tbody>
<tr>
<td>$E(\log T_{11})$</td>
<td>4.872</td>
<td>0.268</td>
<td>0.246</td>
<td>(4.347, 5.397)</td>
<td>(4.390, 5.354)</td>
</tr>
<tr>
<td>$E(\log T_{12})$</td>
<td>3.979</td>
<td>0.216</td>
<td>0.243</td>
<td>(3.556, 4.402)</td>
<td>(3.503, 4.455)</td>
</tr>
<tr>
<td>$sd(\log T_{11})$</td>
<td>1.898</td>
<td>0.359</td>
<td>0.199</td>
<td>(1.194, 2.602)</td>
<td>(1.508, 2.288)</td>
</tr>
<tr>
<td>$sd(\log T_{12})$</td>
<td>2.170</td>
<td>0.188</td>
<td>0.294</td>
<td>(1.802, 2.538)</td>
<td>(1.594, 2.746)</td>
</tr>
<tr>
<td>$\rho$</td>
<td>0.232</td>
<td>0.107</td>
<td>0.114</td>
<td>(0.022, 0.442)</td>
<td>(0.009, 0.455)</td>
</tr>
<tr>
<td>$S(20, 20)$</td>
<td>0.583</td>
<td>0.035</td>
<td>0.045</td>
<td>(0.514, 0.652)</td>
<td>(0.495, 0.671)</td>
</tr>
<tr>
<td>$S(30, 30)$</td>
<td>0.501</td>
<td>0.036</td>
<td>0.048</td>
<td>(0.430, 0.572)</td>
<td>(0.407, 0.595)</td>
</tr>
<tr>
<td>$S(40, 40)$</td>
<td>0.446</td>
<td>0.035</td>
<td>0.050</td>
<td>(0.377, 0.545)</td>
<td>(0.348, 0.544)</td>
</tr>
<tr>
<td>$S(50, 50)$</td>
<td>0.405</td>
<td>0.035</td>
<td>0.050</td>
<td>(0.336, 0.474)</td>
<td>(0.307, 0.503)</td>
</tr>
<tr>
<td>$S(60, 60)$</td>
<td>0.373</td>
<td>0.035</td>
<td>0.050</td>
<td>(0.304, 0.442)</td>
<td>(0.275, 0.453)</td>
</tr>
<tr>
<td>$S(70, 70)$</td>
<td>0.346</td>
<td>0.039</td>
<td>0.049</td>
<td>(0.270, 0.422)</td>
<td>(0.250, 0.442)</td>
</tr>
<tr>
<td>$S(80, 80)$</td>
<td>0.322</td>
<td>0.035</td>
<td>0.049</td>
<td>(0.253, 0.391)</td>
<td>(0.226, 0.418)</td>
</tr>
<tr>
<td>$S(150, 150)$</td>
<td>0.216</td>
<td>0.074</td>
<td>0.048</td>
<td>(0.071, 0.361)</td>
<td>(0.122, 0.310)</td>
</tr>
<tr>
<td>$S(360, 360)$</td>
<td>0.103</td>
<td>0.070</td>
<td>0.053</td>
<td>(0.000, 0.240)</td>
<td>(0.000, 0.207)</td>
</tr>
</tbody>
</table>

* is the 95% CI using the estimate analytic standard error; ** is the 95% CI by using Bootstrap standard error based on 100 bootstrap samples.
Figure 6.9: Estimated bivariate density for DRS data
Figure 6.10: (a) Estimated marginal density of $T_{i1}$; (b) Estimated marginal density of $T_{i2}$. 
Figure 6.11: Estimated bivariate survival function $P_r(T_1 \geq t_1, T_2 \geq t_2)$ for DRS data
Figure 6.12: Estimated bivariate survival function $P_r(T_1 \geq t, T_2 \geq t)$ for DRS data
Bibliography


Cook, R.D. and Johnson, M.E. (1981). A family of distributions for modeling non-


Moran, P.A.P. (1967). Testing for correlation between non-negative variates. *Bi-


van der Laan, M.J. (1996). Efficient estimation in the bivariate censoring model and


